

Biomedical Imaging, Vision and Learning Laboratory



Universidad
Industrial de
Santander

**DETECCIÓN DE NÓDULOS PULMONARES USANDO UNA REPRESENTACIÓN
PROFUNDA DE LOCALIZACIÓN EN SECUENCIA DE TOMOGRAFÍA
COMPUTARIZADA (DEEPLL)**

JONATHAN YESID HERNANDEZ BAUTISTA

DYLAN YESID VILLALBA ROA

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICOMECÁNICAS
ESCUELA DE INGENIERÍA DE ELÉCTRICA, ELECTRÓNICA Y DE
TELECOMUNICACIONES
BUCARAMANGA**

2023

**DETECCIÓN DE NÓDULOS PULMONARES USANDO UNA REPRESENTACIÓN
PROFUNDA DE LOCALIZACIÓN EN SECUENCIA DE TOMOGRAFÍA
COMPUTARIZADA (DEEPLL)**

**JONATHAN YESID HERNANDEZ BAUTISTA
DYLAN YESID VILLALBA ROA**

**Trabajo de Grado para optar al título de:
Ingeniero Electrónico**

**Director:
Fabio Martínez Carrillo
Doctor en Ingeniería de sistemas y computación**

**Codirector:
Franklin Alexander Sepúlveda
Doctor en Ingeniería-Automática**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICOMECÁNICAS
ESCUELA DE INGENIERÍA DE ELÉCTRICA, ELECTRÓNICA Y DE
TELECOMUNICACIONES
BUCARAMANGA**

2023

AGRADECIMIENTOS

Principalmente a Dios y a mis padres Gerardo Benitez y Martha Bautista que siempre me han apoyado en todo momento, que siempre me han brindado su ayuda y son mi mas grande motivación para poder cumplir mis sueños y poder brindarle un poquito de todo lo que me han dado.

A mi hermana mayor Liseth Hernandez, que ha estado para mi desde que era pequeño y siempre me ha ayudado motivacionalmente y emocionalmente a siempre seguir adelante y a mis sobrinos que son una de las mas grandes motivaciones que tengo de poder brindarles un futuro mejor.

A mi director de tesis, el Profesor Fabio Martinez, quien con su gran paciencia, dedicación, profesionalismo y vocación, me enseñó una forma de poder cambiar el mundo mediante la investigación y que siempre ha estado ahí a lo largo de este proyecto.

Al grupo de investigación *BivL²ab*, que más que un grupo, son una familia en la cual conocí el mundo de la investigación y siempre me han ayudado a solucionar todas mis inquietudes sobre todo a Alejandra Moreno y Luis guayacán los cuales desde el inicio de este proyecto me han ayudado, apoyado y brindado todo su conocimiento de la forma mas amable posible y con toda la paciencia posible.

A mi compañero de tesis Dylan Villalba con el cual compartí a lo largo de esta carrera y con el que siempre nos hemos apoyado en los momentos que hemos pasado a lo largo de esta.

Y por ultimo y no menos importante a la UIS y a la E3T y a sus profesores como mi codirector Franklin Sepulveda por apoyarme y abrirme las puertas para poder estudiar en esta gran universidad y guiarme por el camino correcto para ser un gran profesional.

Jonathan Hernandez

Hoy cierro un capítulo en mi vida
y aunque no sé qué me deparará el futuro
siento la alegría y la emoción
de haber llegado hasta aquí, a la cima de mi formación.

La UIS me ha brindado
conocimientos, retos y experiencias
académica, políticas, culturales, deportivas y sociales
que me han llevado a conocer mi esencia.

He conocido gente maravillosa
en mi camino he encontrado
amigos como Jonathan Hernández, Paola Oquendo
con quien he vivido momentos memorables.

A mi director Fabio Martínez
por la paciencia y coordinación
en el proyecto me inspiró
a ser mejor cada día.

Alejandra Moreno y Luis Guayacán
también han sido parte de este sendero

Y al codirector Franklin Sepulveda
le agradezco por su apoyo,

También agradezco a cada uno de los miembros
Del grupo de investigación *BivL²ab*
por haber creído en mí y motivarme
en los momentos en los que pensé en desistir.

Agradezco finalmente a mis padres,
Liliana Roa y Wiston Villalba por su incondicional apoyo
por su amor, su paciencia y su sacrificio.
sin ellos no estaría aquí celebrando mi logro con tanto entusiasmo y júbilo.

La universidad se queda atrás
y un nuevo camino se abre ante mí
lleno de incertidumbre y muchos retos
pero también de oportunidades y de alegrías.

Gracias a la universidad
por haberme enseñado tanto
y gracias a la vida
porque sé que aún hay más por aprender y vivir.

Dylan Villalba

CONTENIDO

	pág.
Introducción	12
1. Fundamentos y Trabajo Previo	15
1.1. Cáncer de pulmón	15
1.2. Métodos de aprendizaje profundo para la detección de objetos	16
1.2.1. Faster R-CNN:	18
1.2.2. YOLO:	19
1.3. Modelos de explicabilidad	21
1.3.1. CAM	21
1.3.2. Grad-CAM	22
1.4. Aprendizaje profundo para la detección automática de los nódulos pulmonares	24
2. Problema de Investigación	28
3. Objetivos	30
4. Método Propuesto	31
4.1. Extractor de características con múltiples campos receptivos	31
4.2. Representación de localización multicanal	33
4.3. Explicabilidad por medio de componentes principales	35
4.4. Generación de nódulos candidatos regularizados por atención	36
5. CONFIGURACIÓN EXPERIMENTAL	41
5.1. Configuración del modelo	42

6. EVALUACIÓN Y RESULTADOS	44
7. CONCLUSIONES Y PERSPECTIVAS	52
BIBLIOGRAFÍA	54

LISTA DE FIGURAS

	pág.
Figura 1. Tipos de nódulos en imágenes TC.	17
Figura 2. Descripción general de la arquitectura Faster R-CNN.	19
Figura 3. Descripción general del funcionamiento de la YOLO	20
Figura 4. Obtención de mapas de activación de clases CAM	22
Figura 5. Mapas de activación por el gradiente Grad-CAM	23
Figura 6. Enfoque propuesto basado en la YOLOv5	32
Figura 7. Bloques C3 y SPPF	33
Figura 8. Eigen-CAM	36
Figura 9. Activaciones Eigen-CAM	37
Figura 10 Localización con mapas de saliencia.	39
Figura 11 Imágenes TC del conjunto de datos LIDC-IDRI	42
Figura 12 Curva FROC de los distintos modelos de la YOLO	45
Figura 13 Modelo de barras YOLOv5	46
Figura 14 Reducción de falsos positivos	48
Figura 15 Análisis observacional de los nódulos.	51

LISTA DE TABLAS

	pág.
Tabla 1. YOLOv5s con diferentes tamaños de cuadros de anclaje.	47
Tabla 2. YOLOv5s con diferentes números de cuadros de anclaje.	47
Tabla 3. Sensibilidad obtenida de los mapas de saliencia en diferentes etapas.	48
Tabla 4. Modelo propuesto con reducción de falsos positivos.	49

Resumen

Título: Detección de nódulos pulmonares usando una representación profunda de localización en secuencia de tomografía computarizada (DEEPLL). *

Autor: Jonathan Yesid Hernandez Bautista y Dylan Yesid Villalba Roa. **

Palabras Clave: Tomografía computarizada, aprendizaje profundo, nódulos pulmonares, localización, cáncer.

Descripción: Los nódulos son el principal indicador de cáncer de pulmón, generalmente observados como masas anormales en estudios de tomografía computarizada (TC). Sin embargo, en la rutina clínica se reporta una baja y muy variable sensibilidad, reportando resultados que van desde el 30 % al 97 %. Esta variación en la tarea de detección puede estar asociada a la variabilidad textural de los nódulos y a la subjetividad inherente a las observaciones de los expertos. A pesar de los múltiples enfoques para apoyar la detección de nódulos, la caracterización respectiva sigue siendo desafiante, especialmente para masas diminutas con sólo un 51 % de sensibilidad. Este trabajo introduce una representación profunda YOLO multicanal que tiene la capacidad de localizar nódulos, representados a múltiples escalas en rutas de procesamiento independientes. El enfoque propuesto recupera cuadros delimitadores de diferentes rutas de procesamiento, localizando nódulos que pueden variar drásticamente en su tamaño. Además, en este trabajo se recuperan mapas de explicabilidad a partir de una eigen descomposición, que permite comparar las predicciones y reducir los falsos positivos. El enfoque propuesto se validó con respecto al conjunto de datos público LIDC, tomando nódulos con una mediana de acuerdo entre al menos tres radiólogos. En dicho conjunto de datos, el enfoque propuesto alcanzó un 88.01 % de métrica de rendimiento competitivo (por sus siglas en inglés: CPM), y una sensibilidad del 93.64 %.

* Trabajo de grado

** Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería de Eléctrica, Electrónica y de Telecomunicaciones. Director: Fabio Martínez Carrillo, Ph.D., Codirector: Franklin Alexander Sepúlveda, Ph.D

Abstract

Title: Detection of pulmonary nodules using deep localization rendering in computed tomography sequences (DEEPLL) *

Author: Jonathan Yesid Hernandez Bautista y Dylan Yesid Villalba Roa **

Keywords: Computed tomography, deep learning, pulmonary nodules, localization, cancer

Description: Nodules are the main indicator of lung cancer, generally observed as abnormal masses in computed tomography (CT) studies. However, in routine clinical practice, low and highly variable sensitivity is reported, with results ranging from 30 % to 97 %. This variation in the detection task may be associated with the textural variability of nodules and the inherent subjectivity of expert observations. Despite multiple approaches to support nodule detection, the respective characterization remains challenging, especially for tiny masses with only 51 % sensitivity. This work introduces a deep YOLO multichannel representation that has the ability to locate nodules, represented at multiple scales in independent processing paths. The proposed approach retrieves bounding boxes from different processing paths, locating nodules that can vary drastically in size. In addition, this work recovers explanatory maps from an eigen decomposition, which makes it possible to compare predictions and reduce false positives. The proposed approach was validated against the public LIDC dataset, taking nodules with a median agreement between at least three radiologists. In that dataset, the proposed approach achieved 88.01 % of competition performance metric (CPM), and a sensitivity of 93.64 %.

* Degree work

** School of Physical-Mechanical Engineering. School of Electrical, Electronic and Telecommunications Engineering. Advisor: Fabio Martínez Carrillo Ph.D., Co-Advisor: Franklin Alexander Sepúlveda, Ph.D

Introducción

El cáncer de pulmón (CP) es el segundo cáncer más prevalente a nivel mundial, reportando cerca de 2,2 millones de casos, superado por el cáncer de mama. Además, el CP registra una elevada tasa de mortalidad, con al menos 1,8 millones de muertes relacionadas en 2020, teniendo una tasa de supervivencia a 5 años de sólo el 21 %¹. Los nódulos pulmonares son el principal biomarcador para diagnosticar el CP. Estos nódulos son masas tisulares y varían de [3-30]mm, presentando fuertes variaciones en cuanto a su tamaño, localización y representación de la apariencia². Los nódulos suelen observarse en estudios de tomografía computarizada (TC), pero su análisis representa un reto considerable para los radiólogos debido a las estructuras del parénquima pulmonar y los vasos sanguíneos. De hecho, la detección de nódulos en estas secuencias reportan una baja sensibilidad, especialmente en estadios tempranos se estima que un 25 % de los nódulos son pasados por alto por los radiólogos durante los exámenes de TC³. Además, esta tarea es altamente subjetiva y dependiente de la experiencia del radiólogo, lo que exige un apoyo adicional en factores clínicos como la edad, la localización de otros hallazgos anormales y los

¹ SUNG, Hyuna, *et al.* "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries". En: *CA: a cancer journal for clinicians* 71.3 (2021), págs. 209-249.

² MONKAM, Patrice, *et al.* "Detection and Classification of Pulmonary Nodules Using Convolutional Neural Networks: A Survey". En: *IEEE Access* 7 (2019), págs. 78075-78091. DOI: 10.1109/ACCESS.2019.2920980.

³ DUMA, Narjust; SANTANA-DAVILA, Rafael y MOLINA, Julian R. "Non-small cell lung cancer: epidemiology, screening, diagnosis, and treatment". En: *Mayo Clinic Proceedings*. Vol. 94. 8. Elsevier. 2019, págs. 1623-1640.

informes complementarios de la historia clínica ⁴.

En la literatura han surgido enfoques computacionales para apoyar la localización y caracterización de nódulos. Las estrategias preliminares se han dedicado a construir descriptores, explotando características geométricas como circularidad, redondez, excentricidad o elipticidad ⁵. Recientemente, la *Faster R-CNN* ha sido ajustada para la localización de nódulos aprovechando su red de propuesta regional (por sus siglas en inglés: RPN), generando regiones de interés (por sus siglas en inglés: ROI) a través de las diferentes escalas. Adicionalmente, esta estrategia se ha extendido para realizar segmentación utilizando para ello una etapa previa de segmentación mediante capas convolucionales que se encargan de extraer las características de la imagen de entrada ^{6 7}. También, la arquitectura YOLO se ha adaptado para localizar nódulos, prediciendo cuadros relacionados con una probabilidad de nódulo de clase en un espacio de cuadrícula de imagen. Por ejemplo, utilizando la versión YOLOv3, con una representación de convolución que aplica múltiples escalas de patrones de características de nódulos que se caracterizan por su extractor automático de características multiescala y su generador de cajas delimitadoras ⁸. Asimismo, se ha adaptado una estrategia de red en U (por su nombre en inglés: U-Net) mejora-

⁴ DEL CIELLO, Annemilia, *et al.* "Missed lung cancer: when, where, and why?" En: *Diagnostic and interventional radiology* 23.2 (2017), pág. 118.

⁵ RUBIN, Geoffrey D. "Lung nodule and cancer detection in CT screening". En: *Journal of thoracic imaging* 30.2 (2015), pág. 130.

⁶ LIU, Kai, *et al.* "Evaluating a fully automated pulmonary nodule detection approach and its impact on radiologist performance". En: *Radiology: Artificial Intelligence* 1.3 (2019), e180084.

⁷ HAN, Hao, *et al.* "Fast and adaptive detection of pulmonary nodules in thoracic CT images using a hierarchical vector quantization scheme". En: *IEEE journal of biomedical and health informatics* 19.2 (2014), págs. 648-659.

⁸ LIU, Chenyang, *et al.* "Automatic detection of pulmonary nodules on CT images with YOLOv3: development and evaluation using simulated and patient data". En: *Quantitative Imaging in Medicine and Surgery* 10.10 (2020), pág. 1917.

da a regiones volumétricas de TC de entrada, que a su vez adapta características extraídas de una Faster R-CNN para la detección de nódulos ⁹. Sin embargo, estas arquitecturas presentan una baja sensibilidad para objetos pequeños, generando un número notable de falsos positivos.

En este trabajo de grado se presenta una adaptación de la arquitectura YOLO para la localización de nódulos. Esta arquitectura YOLO se modela en términos de múltiples canales y una representación de múltiples escalas convolucionales para enriquecer la representación de las imágenes en TC. La arquitectura propuesta es capaz de localizar y detectar nódulos con notable sensibilidad, incluso para masas pequeñas. El enfoque propuesto utiliza múltiples ramas espacial convolucional que proporcionan mapas de saliencia enriquecidos para encontrar nódulos a diferentes escalas. A continuación, a partir de dichos mapas espaciales, se recuperan fotogramas de anclaje que se adaptan a la localización de nódulos pulmonares. Por último, se adapta una Eigen-CAM para explicar la detección de nódulos y proporcionar un esquema de supervisión alternativo para alertar de posibles nódulos. Estos mapas Eigen-CAM permiten ser operados en paralelo desde capas convolucionales y servir como filtros de activaciones nodulares falsas, es decir, reduciendo el número de falsos positivos.

⁹ SHI, Lukui; MA, Hongqi y ZHANG, Jun. "Automatic detection of pulmonary nodules in CT images based on 3D Res-1 network". En: *The Visual Computer* 37 (2021), págs. 1343-1356.

1. Fundamentos y Trabajo Previo

En esta sección se presenta una descripción de los fundamentos teóricos relacionados al diagnóstico del cáncer pulmonar. En segundo lugar, se revisan algunos métodos de aprendizaje profundo para la detección automática de objetos, aplicables a la detección de nódulos pulmonares. Después, se realizará un resumen de los métodos de explicabilidad y finalmente un breve repaso al estado del arte sobre trabajos previos para la localización de lesiones pulmonares.

1.1. Cáncer de pulmón

El principal biomarcador del cáncer de pulmón son los nódulos pulmonares, masas con un diámetro que puede variar entre 3 a 30 mm. Hasta el 40 % de los nódulos pulmonares pueden ser malignos especialmente los que tienen un diámetro mayor a 10 mm. Las características morfológicas útiles para detectar y estimar la malignidad de los nódulos se describen a continuación y se muestran en la Figura 1.

- **Tamaño.** Es una de las principales características para reconocer la malignidad de un nódulo. Al tener un mayor tamaño en su diámetro el nódulo aumenta su posibilidad de malignidad, teniendo menos del 1 % para nódulos menores de 6 mm (como los micro-nódulos los cuales generalmente no miden más de 3 mm), entre 0.5 % y 2 % para aquellos entre 6 a 8 mm y, aproximadamente 3 % para mayores de 8 mm.
- **Atenuación.** Es una característica relacionada con la densidad textural del tejido, que puede clasificarse como:
 - Sólido: áreas con quistes o líquidas, además de ser los mas comunes por ser densos y homogéneos.

- Vidrio esmerilado: es un área que tiene aumento de la atenuación focal que no oculta ni impide ver los vasos subyacentes, su densidad es mayor a la del parénquima pulmonar y en ocasiones puede rodear también los nódulos sólidos.
 - Sub-sólido ó Parcialmente solidos: opacidades de vidrio esmerilado puro. Además, ocurren con menos frecuencia que los nódulos sólidos, pero tienen una tasa de malignidad mucho mayor y la mayoría de estos cuando se detectan suelen ser lesiones precancerosas o en estado inicial de cáncer de pulmón.
- **Localización.** el cáncer pulmonar es frecuentemente identificado en lóbulos superiores. Según su localización estos se pueden clasificar como:
- Yuxtopleural: son pequeños nódulos solidos adyacentes a las fisuras o a la superficie pleural y son usualmente benignos.
 - Yuxtavascular: son nódulos adheridos a los vasos sanguíneos. Son difíciles de detectar por tener un tono de gris similar a la de los vasos.
 - Aislados: está rodeado por tejido pulmonar y no está asociado con ninguna otra anomalía en el pulmón o en los ganglios linfáticos cercanos. La mayoría de los nódulos aislados son benignos.

1.2. Métodos de aprendizaje profundo para la detección de objetos

El diagnóstico asistido por computadora (por sus siglas en inglés: CAD) ha sido de gran ayuda para un diagnóstico temprano del cáncer de pulmón. Particularmente en detección de nódulos, los sistemas CAD suelen constar de cinco pasos básicos como lo son: adquisición de datos, pre-procesamiento, segmentación pulmonar, detección de nódulos y reducción de falsos positivos. La segmentación pulmonar tiene

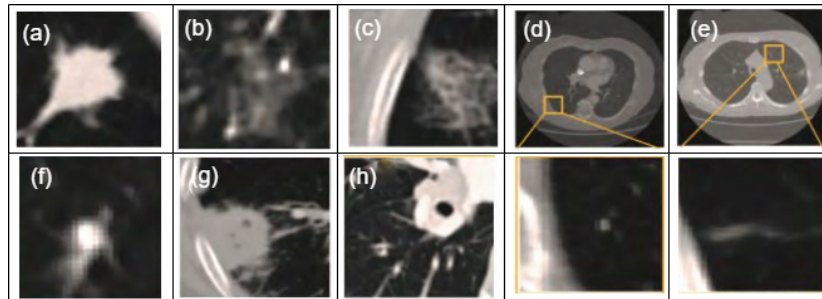


Figura 1. Tipos de nódulos en imágenes TC. (a) Nódulo sólido; (b) Nódulo parcialmente sólido; (c) Nódulo vidrio esmerilado; (d) Micro-nódulo; (e) No-nódulo; (f) Nódulo aislado; (g) Nódulo yuxtopleural; (h) Nódulo yustavascular.

como objetivo principal extraer el parénquima pulmonar de otros órganos y tejidos, hacer esto de forma exitosa ayuda en gran medida la detección de los nódulos. En la detección de nódulos se busca separar todos los nódulos sospechosos en un área de interés para que el sistema CAD pueda funcionar mejor con una alta sensibilidad. Por último, en la reducción de los falsos positivos se centra en clasificar los nódulos pulmonares verdaderos de los nódulos sospechosos y así eliminar los nódulos falsos reduciendo la tasa de falsos positivos en el sistema CAD¹⁰.

En los últimos años, las técnicas de aprendizaje automático han presentado importantes avances en la detección de objetos. Particularmente, las redes neuronales artificiales (RNAs), las cuales se basan en funciones no lineales, organizadas jerárquicamente para representar la relación de las características de entrada y sus correspondientes etiquetas. En cuanto a la representación de imágenes, las arquitecturas convolucionales (por sus siglas en inglés: CNN) codifican patrones visuales y brindan optimización en tareas relacionadas a las clasificaciones de imágenes¹¹. A continuación, se presentan dos de las arquitecturas que mejor desempeño han

¹⁰ ZHANG, Guobin, *et al.* "Automatic nodule detection for lung cancer in CT images: A review". En: *Computers in biology and medicine* 103 (2018), págs. 287-300.

¹¹ FAN, Jianqing; MA, Cong y ZHONG, Yiqiao. "A selective overview of deep learning". En: *Statistical science: a review journal of the Institute of Mathematical Statistics* 36.2 (2021), pág. 264.

mostrado en la tarea de detección.

1.2.1. Faster R-CNN: La red neuronal convolucional basada en regiones (por sus siglas en inglés: R-CNN) fue una de las primeras arquitecturas diseñadas para tareas de detección en imágenes naturales. La base de esta arquitectura funciona generando propuestas de regiones candidatas a contener un objeto, a través de la búsqueda selectiva. Cada una de las regiones candidatas es ingresada a una CNN con la finalidad de extraer características de la imagen para, posteriormente, alimentar una máquina de soporte vectorial (por sus siglas en inglés: SVM). Sin embargo, el tiempo de inferencia de estas redes es considerablemente grande y el algoritmo de búsqueda selectiva puede presentar diversos falsos positivos en su generación de candidatos.

Más recientemente, fue introducida la arquitectura Faster R-CNN que hace uso de la red de propuestas regionales (por sus siglas en inglés: RPN), permitiendo generar regiones de interés, por medio de cuadros de anclaje a diferentes escalas y proporciones. La Faster R-CNN (Figura 2) a diferencia de otras técnicas de detección de objetos, realiza una etapa previa de segmentación utilizando capas convolucionales, las cuales extraen los mapas de características de la imagen de entrada. Estos mapas son ingresados como entrada a la RPN, la cual utiliza una ventana deslizante sobre estas características, generando k cuadros de anclaje con diferentes formas y tamaños. A partir del conjunto de cuadros de anclaje extraídos se estima, tanto la probabilidad de que cada uno contenga un objeto, así como el cuadro de anclaje con mayor probabilidad de contenerlo. Los cuadros delimitadores resultantes pasan a la capa de agrupación de regiones de interés (RoI) en donde a partir de cada propuesta se extraen mapas de características de tamaño fijo para cada cuadro de anclaje. Finalmente, estos mapas de características pasan por una capa completamente conectada (*Fully-connected*) y una capa de regresión lineal que clasificará el objeto y delimitará el cuadro identificado, respectivamente.

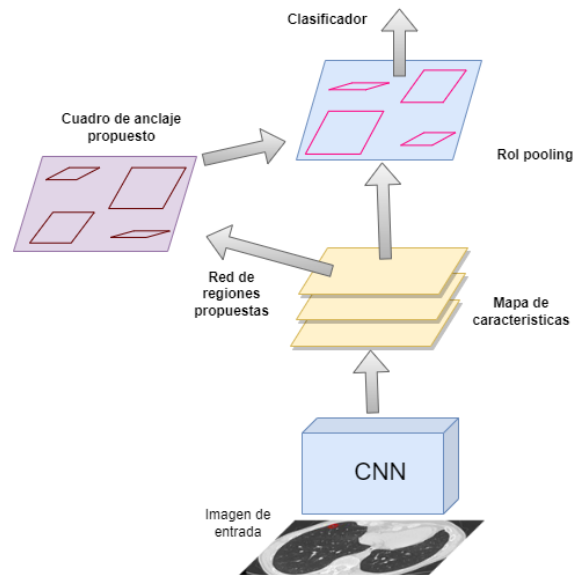


Figura 2. Descripción general de la arquitectura Faster R-CNN.

1.2.2. YOLO: La arquitectura YOLO (por sus siglas en inglés: *You Only Look Once*), utiliza una red convolucional *end-to-end* (entrenamiento de extremo a extremo de la arquitectura) para la detección y localización de objetos en tiempo real. Esta arquitectura primero realiza una segmentación de la imagen a partir de unas más pequeñas, dividiéndola en una grilla cuadrada de dimensiones $S \times S$ en donde cada una presenta una tarea diferente en la detección. Para cada celda de la grilla se definen B cuadros delimitadores con una puntuación de confianza C definida de la siguiente forma: $C = PR(X) * IOU(Y', Y)$ en donde Y' es el valor predicho; Y es el valor real; e $IOU(Y', Y)$ representa la intersección sobre la unión, la cual describe la superposición entre dos cuadros delimitadores y varía entre 0 y 1 (donde 1 corresponde a un solapamiento total).

Por último, $PR(X)$ representa la probabilidad de que haya un objeto de interés en el cuadro delimitador B . La predicción del cuadro delimitador B_i de una celda está compuesto por 5 elementos que son: (x, y, w, h, c) , en donde (x, y) representan las

coordenadas del centro del cuadro delimitador B_i y (w, h) representan el ancho y el alto del mismo, respectivamente. Dado este tensor, la YOLO puede calcular la pérdida y así mismo corregir el centro y el cuadro delimitado de cada predicción en donde se busca una IoU (intersección sobre la unión, una medida de intersección) mayor a un umbral que pueda delimitar al objeto de la mejor forma. Finalmente se une la clasificación de los objetos y así mismo la predicción B de la manera $S \times S \times B \times (5 + No.Clases)$ la cual será la entrada del algoritmo de supresión no máxima que se encargará de elegir el cuadro delimitador con el mayor nivel de confianza y de descartar los cuadros delimitadores con un nivel de confianza menor a un umbral y con un valor alto de IOU y así obtener el cuadro que mejor delimite al objeto de interés ^{12 13}. En la Figura 3 se puede apreciar gráficamente el funcionamiento de la arquitectura YOLO.

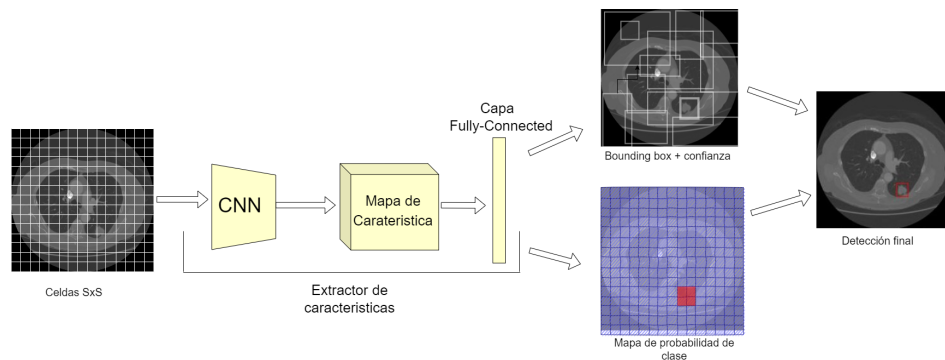


Figura 3. Descripción general del funcionamiento de la YOLO

¹² WU, Wentong, *et al.* "Application of local fully Convolutional Neural Network combined with YOLO v5 algorithm in small target detection of remote sensing image". En: *PloS one* 16.10 (2021), e0259283.

¹³ HUANG, Rachel; PEDOEEM, Jonathan y CHEN, Cuixian. "YOLO-LITE: a real-time object detection algorithm optimized for non-GPU computers". En: *2018 IEEE international conference on big data (big data)*. IEEE. 2018, págs. 2503-2510.

1.3. Modelos de explicabilidad

Las técnicas de aprendizaje profundo han demostrado un desempeño excepcional en tareas de reconocimiento visual, como la clasificación de imágenes y la detección de objetos. Sin embargo, este tipo de trabajos no reportan de manera clara como la arquitectura obtiene la predicción por lo que carecen de explicabilidad e interpretabilidad clínica. De hecho, para la transferencia en ambientes clínicos, la explicabilidad de la predicción realizada por la red, representa gran importancia en diferentes entornos, porque es una herramienta de apoyo para respaldar decisiones médicas, como el diagnóstico temprano o el inicio de un tratamiento adecuado.

Existen múltiples métodos que ofrecen una explicación visual de las tareas desempeñadas por los modelos, proporcionando la zona de interés en donde la arquitectura concentra su atención para realizar cierta predicción. La mayoría de estas técnicas utilizan la arquitectura completa de la red, actuando sobre los datos de entrada, las predicciones de salida y las activaciones que se producen en las capas ocultas otorgando la predicción de la clase. La visualización de estas zonas de interés sobre los datos de entrada permite la explicabilidad de los resultados dados por la arquitectura. A continuación, se presenta brevemente algunos de estos métodos.

1.3.1. Mapas de activación de clase (CAM): El mapa de activación de clase ¹⁴ (por sus siglas en inglés: CAM) es una estrategia que permite generar mapas de saliencia a través de las regiones de los mapas de activaciones que presentan un mayor aporte/contribución con respecto a la predicción final de una clase. Para esto se realiza un “*Global Average Pooling*” (GAP) permitiendo generar un promedio global para cada mapa de activación y la concatenación de cada uno de ellos ge-

¹⁴ ZHOU, Bolei, *et al.* “Learning deep features for discriminative localization”. En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, págs. 2921-2929.

nerando un vector resultante el cual es la entrada de la capa softmax para realizar la clasificación. Esto permite obtener una correspondencia mayor entre los mapas de características y las clases. Una ventaja del GAP es que no utiliza un parámetro de aprendizaje. Además de que la agrupación de los promedios globales suma la información espacial de los mapas de activación, permitiendo que sea más robusta a los cambios de traslación de la entrada.

Matemáticamente, considerando $f_k(x, y)$ representa la activación de la última capa convolución en la ubicación espacial (x, y) , se realiza la ponderación de promedios global sobre la última capa convolucional quedando de la siguiente forma $F^k = \sum_{x,y} f_k(x, y)$, finalmente para cada clase c , la salida de la *softmax* viene dada por $S_c = \sum_k w_k^c f_k(x, y)$, en donde cada clase obtiene un peso correspondiente denotado como w_k^c . Se puede visualizar el método en la Figura 5.

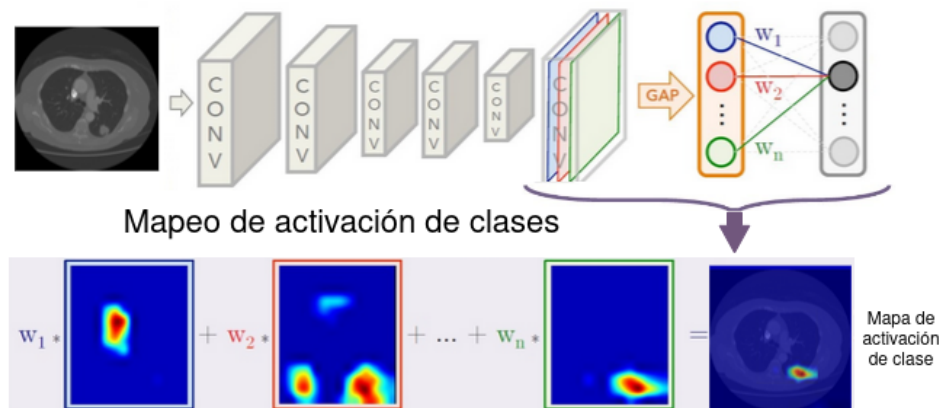


Figura 4. Método para obtención de mapas de activación de clases CAM propuesto en ¹⁵.

1.3.2. Mapas de activaciones de clase ponderados por gradientes (Grad-CAM):

Los mapas de activaciones de clase ponderados por gradientes (Grad-CAM) ¹⁶ genera mapas visuales con respecto a la decisión de la red para cada predicción de

¹⁶ SELVARAJU, Ramprasaath R, *et al.* "Grad-CAM: Why did you say that?" En: *arXiv preprint arXiv:1611.07450* (2016).

clase, sobre cualquier arquitectura convolucional. El método consiste en calcular el gradiente correspondiente a la clase de salida de interés. Con el gradiente obtenido, se realiza la combinación lineal de la capa convolucional, de manera análoga al método CAM.

Los cálculos de los pesos se realizan por medio de una media aritmética de los gradientes de la salida de nuestra clase de interés y , con respecto a las activaciones A^k de la capa convolucional. Los valores de los gradientes son ponderados con la finalidad de obtener un valor que represente la importancia de la neurona a_k^c de la capa convolucional en la decisión final, en donde a_k^c es igual a $a_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$, a_k^c el cual resalta la importancia de los mapas de activaciones k para la clase de interés y , la combinación lineal de los mapas activaciones A_k y los respectivos pesos calculados a_k^c . A estos resultados se les aplica una función de activación ReLU, $L_{GradCAM}^c = ReLU(\sum_k a_k^c A_k)$ ¹⁶.

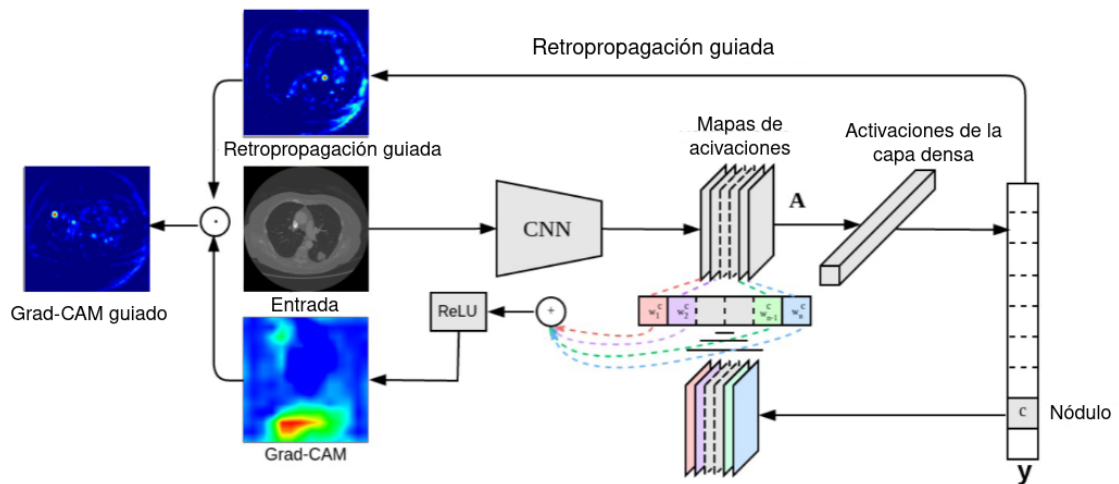


Figura 5. Método para obtención de mapas de activación de clases ponderado por el gradiente Grad-CAM propuesto en ¹⁷.

1.4. Aprendizaje profundo para la detección automática de los nódulos pulmonares

Los sistemas CAD han logrado avances significativos en la localización de nódulos pulmonares. Una de las más utilizadas son las estrategias Faster R-CNN, por ejemplo, Nguyen chi *et. al.* adaptó una agrupación de desplazamiento promedio para aprender configuraciones de cuadros de anclaje, incluyendo también un paso de reducción de falsos positivos ¹⁸. Asimismo, Xie Hongtao *et. al.* propuso una configuración Faster R-CNN para la detección de nódulos con dos redes de propuesta de región y una capa deconvolucional para reducir los candidatos falsos positivos ¹⁹. Del mismo modo, Huang Xia *et. al.* propone localizar parches de nódulos pulmonares, fusionando los candidatos que tienen una distancia euclidiana cercana, e incluyen una reducción de falsos positivos mediante una CNN 2D de tres capas y una segmentación basada en la red totalmente convolucional (por sus siglas en inglés: FCN) ²⁰. Estos métodos siguen siendo limitados en cuanto al número significativo de falsos positivos detectados que, a su vez, se preprocesan lentamente y filtran estas regiones. Otros enfoques han adaptado las arquitecturas U-Net, siguiendo representaciones 3D ²¹, y utilizando convoluciones dilatadas como mecanismo de

¹⁸ NGUYEN, Chi Cuong, *et al.* "Pulmonary nodule detection based on faster R-CNN with adaptive anchor box". En: *Ieee Access* 9 (2021), págs. 154740-154751.

¹⁹ XIE, Hongtao, *et al.* "Automated pulmonary nodule detection in CT images using deep convolutional neural networks". En: *Pattern Recognition* 85 (2019), págs. 109-119.

²⁰ HUANG, Xia, *et al.* "Fast and fully-automated detection and segmentation of pulmonary nodules in thoracic CT scans using deep convolutional neural networks". En: *Computerized Medical Imaging and Graphics* 74 (2019), págs. 25-36.

²¹ CAO, Haichao, *et al.* "A two-stage convolutional neural networks for lung nodule detection". En: *IEEE journal of biomedical and health informatics* 24.7 (2020), págs. 2006-2015.

conjunto, y por supuesto incluyendo etapas de reducción de falsos positivos ²². Para suplir estos enfoques, se dispone de escasa información espacial y baja tasa de detección de micronódulos. Para esta representación espacial, algunos enfoques han incluido redes neuronales de convolución profunda multiescala 3D para la detección de candidatos a nódulos ²³, y una red de propuesta de región (por sus siglas en inglés: RPN) utilizando una estructura tipo U-Net para la detección de candidatos a nódulos y un clasificador 3D DCNN para la reducción de falsos positivos ²⁴. Sin embargo, estos enfoques requieren una cantidad significativa de datos para ajustar las representaciones.

Algunas estrategias han incluido la YOLO como esquema de detección principal, siguiendo una detección local sobre una cuadrícula $S \times S$. Por ejemplo, Liu Kehong *et. al.* adaptó la YOLOv5s, añadiendo agrupación estocástica en la red de agrupación de pirámides espaciales (por sus siglas en inglés: SPP) para mejorar la precisión del reconocimiento ²⁵. Asimismo, Mei Sen *et. al.* implementó un YOLOv4 pero integrando un módulo de atención para mejorar la detección y también incluyó una función de pérdida focal para hacer frente a los lotes de entrenamiento de equilibrio

²² AGNES, S Akila; ANITHA, J y SOLOMON, A Arun. "Two-stage lung nodule detection framework using enhanced UNet and convolutional LSTM networks in CT images". En: *Computers in Biology and Medicine* 149 (2022), pág. 106059.

²³ PENG, Haixin; SUN, Huacong y GUO, Yanfei. "3D multi-scale deep convolutional neural networks for pulmonary nodule detection". En: *Plos one* 16.1 (2021), e0244406.

²⁴ GONG, Li, *et al.* "Automated pulmonary nodule detection in CT images using 3D deep squeeze-and-excitation networks". En: *International journal of computer assisted radiology and surgery* 14 (2019), págs. 1969-1979.

²⁵ LIU, Kehong. "STBi-YOLO: A Real-Time Object Detection Method for Lung Nodule Recognition". En: *IEEE Access* 10 (2022), págs. 75385-75394.

²⁶. En la misma línea, Dong Xiao *et. al.* incluyó un mecanismo de atención, pero sobre la arquitectura YOLOv5 ²⁷. Además, Liu Chenyang *et. al.* personalizó la YOLOv3 adaptando una tarea adicional para estimar el diámetro de los nódulos mediante un generador de cuadros delimitadores ⁸. Sin embargo, estos enfoques siguen siendo limitados en la detección de objetos pequeños y el ajuste del mecanismo de atención requiere una cantidad significativa de datos, que en escenarios clínicos resultan inviables.

Además, una de las principales preocupaciones de estas estrategias propuestas es la escasa explicabilidad de la selección de posibles nódulos y el mecanismo para filtrar los nódulos positivos. De hecho, hasta donde sabemos, actualmente no existen arquitecturas que utilicen la explicabilidad en la detección de nódulos pulmonares mediante métodos supervisados. En general, estos esquemas han sido propuestos para explicar de forma general afectaciones de enfermedades pulmonares sobre redes convolucionales clásicas ²⁸. También, algunos enfoques han sido propuestos para destacar patrones de nódulos sobre radiografías de tórax y apoyar el diagnóstico médico ²⁹. Otros esquemas de explicabilidad han guiado el entrenamiento a partir de mapas de activaciones, siguiendo estrategias débilmente supervisadas. En tal caso, la arquitectura se utiliza para clasificar cortes y, por lo tanto, los mapas

²⁶ MEI, Sen; JIANG, HuiQin y MA, Ling. “YOLO-lung: a practical detector based on improved YOLOv4 for pulmonary nodule detection”. En: *2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE. 2021, págs. 1-6.

²⁷ DONG, Xiao, *et al.* “An improved YOLOv5 network for lung nodule detection”. En: *2021 International Conference on Electronic Information Engineering and Computer Science (EIECS)*. IEEE. 2021, págs. 733-736.

²⁸ GASCOIGNE-BURNS, James y KATSIKIANNIS, Stamos. “A Localisation Study of Deep Learning Models for Chest X-ray Image Classification”. En: *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE. 2022, págs. 01-05.

²⁹ BUSH, Isabel. “Lung nodule detection and classification”. En: *Rep. Stanf. Comput. Sci* 20 (2016), págs. 196-209.

recuperados de tales redes se utilizan como entrada para proceder con un esquema de localización de nódulos ³⁰.

³⁰ FENG, Xinyang, *et al.* "Discriminative localization in CNNs for weakly-supervised segmentation of pulmonary nodules". En: *Medical Image Computing and Computer Assisted Intervention- MIC-CAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III* 20. Springer. 2017, págs. 568-576.

2. Problema de Investigación

El cáncer de pulmón tiene los índices mas altos de mortalidad a nivel mundial, reportando 1.8 millones de muertes en el año 2020, afectando mayormente a los países industrializados. Su principal biomarcador son los nódulos pulmonares, masas de tejido de entre 3 y 30 mm. Por lo tanto, la detección temprana de los nódulos pulmonares es fundamental para reducir las cifras de mortalidad de esta enfermedad ¹. Sin embargo, la detección de los nódulos representa un gran desafío para los médicos, ya que, en etapas tempranas, pueden ser pasados por alto durante los exámenes radiológicos y terminar desarrollándose en un cáncer pulmonar. De hecho, se estima que alrededor del 25 % de los nódulos son pasados por alto por los radiólogos durante los exámenes de TC. En realidad, esta tarea resulta altamente subjetiva y dependiente de la experiencia del radiólogo para evaluar las secuencias tomográficas y correlacionarlas con factores como la edad del paciente, la presencia de otras anomalías y la historia clínica ⁴. Por otra parte, el gran parecido que tienen los nódulos pulmonares con el tejido del parénquima circundante con una densidad ligeramente inferior a la de los nódulos o su ubicación en regiones con grandes vasos, dificulta fuertemente su localización ³.

En la literatura han emergido alternativas computacionales para soportar el análisis cuantitativo y la caracterización de los nódulos pulmonares. Sin embargo, estos sistemas suelen requerir de métodos adicionales de preprocesamiento, exhaustivas fases para la adecuación de regiones de interés, y en general no asocian a la predicción establecida un mecanismo de explicabilidad. Adicionalmente, estos métodos presentan un gran número de falsos positivos, por lo que requieren de un modelo adicional encargado de determinar si los nódulos detectados son realmente nódulos.

los o corresponde a otros tipos de tejidos^{31 10}. Lo anterior limita su aplicabilidad en escenarios reales, siendo hoy en día un problema abierto y en continua exploración.

³¹ ALBERT, Ross H y RUSSELL, John J. "Evaluation of the solitary pulmonary nodule". En: *American family physician* 80.8 (2009), págs. 827-831.

3. Objetivos

Objetivo general

Diseñar e implementar una estrategia de aprendizaje profundo para la localización de nódulos pulmonares en secuencias de tomografía computarizada.

Objetivos específicos

- Diseñar una metodología para la selección y preprocesamiento de estudios de tomografías computarizadas con sospecha de cáncer, incluyendo criterios y anotaciones de múltiples observadores.
- Implementar un modelo de aprendizaje profundo dedicado a localización de nódulos pulmonares sobre estudios de tomografía computarizada.
- Diseñar una estrategia para el entrenamiento y ajuste de la representación profunda dedicada a la localización de los nódulos.
- Implementar una estrategia de explicabilidad sobre la representación de localización para generar mapas de atención sobre las regiones que tuvieron mayor aporte en la probabilidad de salida.
- Validar la estrategia propuesta en cuanto a su capacidad para localizar nódulos pulmonares en secuencias de tomografía.

4. Método Propuesto

Este trabajo introduce un esquema para la detección de nódulos basado en un esquema YOLOv5. El enfoque propuesto comienza codificando varios bloques convolucionales, bajo un esquema multiescala, que enriquecen la representación de los nódulos, a partir de un banco de filtros, pero sin perder la representación de la grilla espacial de referencia. A continuación, se codifican múltiples ramas para proponer múltiples regiones a diferentes escalas, pero enfocándose en pequeñas regiones, que potencialmente representan los nódulos. En una etapa de salida se seleccionan cuadros delimitadores a partir de un mapa de probabilidad de clase, considerando los nódulos con mayor confianza. En este trabajo también exploramos un método alternativo para recuperar mapas explicables a partir de características convolucionales que realzan los nódulos potenciales. Estos mapas apoyan el análisis observacional de los expertos, pero también permiten regularizar los candidatos de salida, reduciendo las regiones de falsos positivos. El proceso general de implementación se ilustra en la Figura 6.

4.1. Extractor de características con múltiples campos receptivos

El enfoque propuesto debe preservar la información espacial, con respecto a la información de los nódulos obtenida dentro de las grillas. Dada una imagen de entrada $I(x) \in R^{H \times H}$ se traza una grilla (S) la cual sirve como un bloque de referencia espacial de las regiones de la imagen. A continuación, la imagen de entrada es proyectada por medio de un extractor de características el cual está compuesto por bloques convolucionales (Fig. 7).

Estas características convolucionales preservan la información espacial para definir posibles zonas donde se puede encontrar el nódulo. A demás de ello se usa un cue-

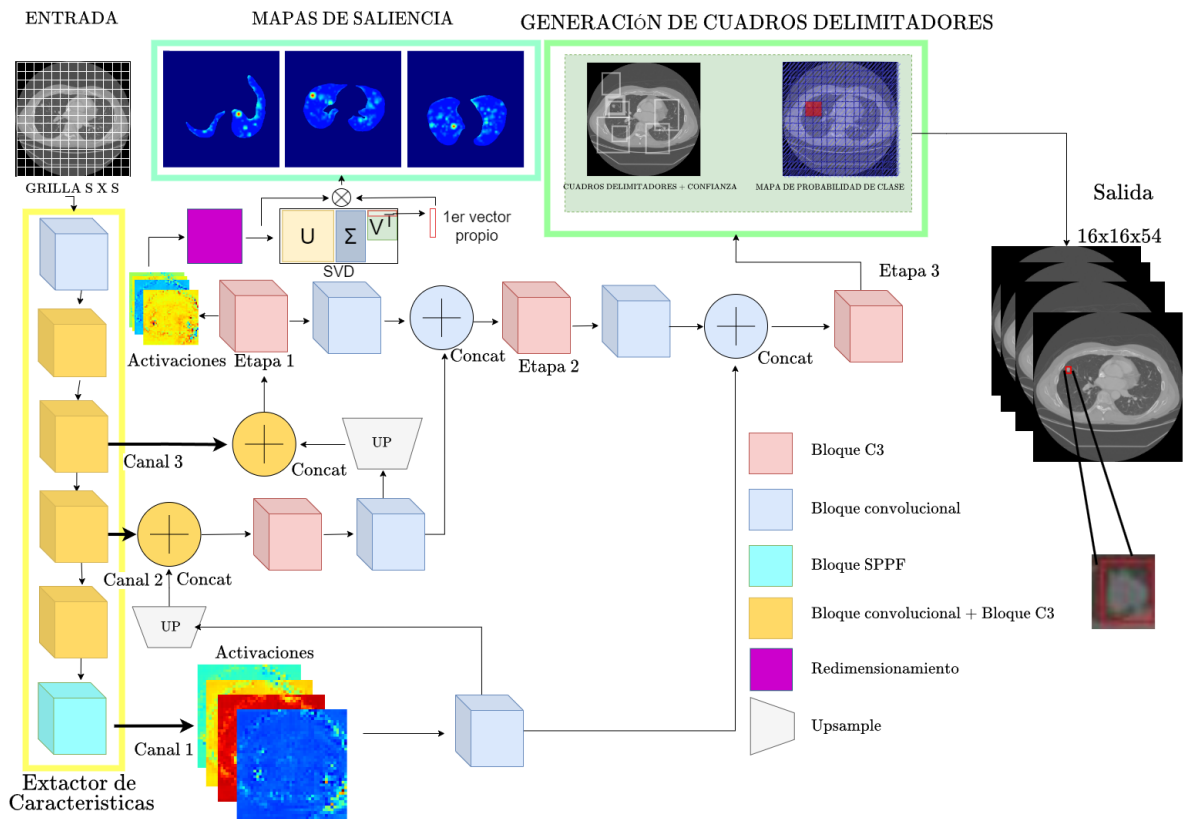


Figura 6. Enfoque propuesto basado en la YOLOv5, donde se visualiza la imagen de entrada con una grilla $S \times S$, el Backbone encargado de extraer las características, las activaciones obtenidas, sus mapas de saliencia y su salida con la generación de los cuadros delimitados.

llo de botella (por su nombre en inglés: *bottleneck*) que permite obtener una mejor cuantificación de las características adicionales logrando un costo computacional bajo y evitando la dispersión del gradiente durante el entrenamiento, teniendo en cuenta que el conjunto de datos de los nódulos suelen ser pequeños, al igual que sus anotaciones etiquetadas, por lo tanto, se necesita un esquema adecuado para el entrenamiento. En este caso existen bloques convolucionales que adicionalmente incluyen un bloque SPPF (Fig. 7) el cual se encarga de obtener un mejor campo receptivo que se utiliza para analizar regiones espaciales alrededor del nódulo en diferentes resoluciones, junto con características propias de este, siguiendo las ope-

raciones de salida de agrupación máxima (por su nombre en inglés: *max pooling*) concatenadas con convoluciones.

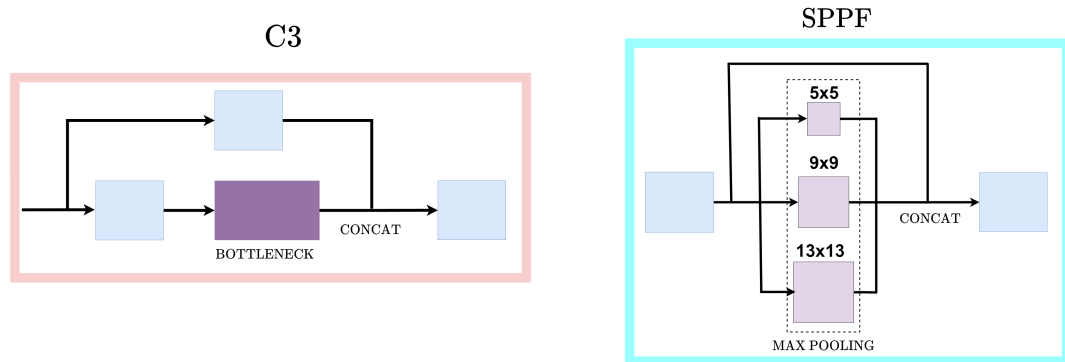


Figura 7. Los bloques convolucionales C3 integran una representación de cuello de botella y los bloques SPPF siguen operaciones de salida de agrupación máxima.

4.2. Representación de localización multicanal

Desde la representación convolucional, construida con múltiples campos receptivos, se generan diferentes canales de procesamiento. Estas ramas de procesamiento capturan información convolucional en diferentes etapas de la representación, teniendo así candidatos que mantienen información con un nivel semántico diferente.

- Canal 1:** Este canal de representación tiene el mayor contenido semántico, extrayéndose directamente del módulo SPPF. En este canal se toman representaciones a diferentes escalas, que contienen información depurada con respecto a las características salientes que tienen mayor presentación en la tarea de entrenamiento, es decir, la detección de nódulos. En este canal de información entonces se toman las activaciones, que proyecta a un bloque convolucional. Desde este bloque convolucional, la información se proyecta en paralelo a la representación final para tomar los nódulos candidatos y al canal 2 de procesamiento. Para poder unificar la información con el canal 2, se lleva a cabo un proceso de sobre muestreo (*upsampling*).

- **Canal 2:** El segundo canal de información es una representación intermedia convolucional. En este caso se tiene una representación semántica intermedia, donde pueden predominar aún características de bajo nivel que representan los datos de entrada. Este bloque convolucional intermedio es concatenado con la representación del canal 2 y proyectado a una representación que utiliza elementos convolucionales organizados como cuello de botella, para extraer información convolucional conjunta de los dos hilos de información, pero preservando la eficiencia es su cálculo (bloque C3). Esta representación conjunta desde C3 es entonces convolucionada y desde ahí la información sirve para reforzar la representación del canal 1, pero también se agrega en una capa final para obtener la localización de los nódulos en la regeneración de regiones de interés (generación de cuadros delimitadores). La proyección agregada hacia el canal 1 también tiene un proceso de sobre-muestreo.
- **Canal 3:** Este canal tiene una representación convolucional primera, con un banco de características que corresponden a un bajo nivel de la imagen. En este caso, existe redundancia en la información, con mayor descripción de la imagen. Este bloque convolucional es concatenado con la información que vienen del canal 2. La información en este bloque es importante en el método propuesto, teniendo en cuenta que luego se realiza un proceso no supervisado para obtener mapas de saliencia que sean complementarios con el mayor nivel semántico aprendido en la tarea supervisada.

Una vez extraída la información y fusionada desde los canales superiores a canales de información cruda (canal 1 y canal 2), la información se procesa en conjunto, obteniendo así una red con características piramidales (por su nombre en inglés: Feature Pyramid Network(FPN)). El bloque conjunto y final de representación permite ser la base para la generación de regiones de interés y clasificación de nódulos.

4.3. Explicabilidad por medio de componentes principales

Para completar la representación de los nódulos, en este trabajo se calcularon los mapas de explicabilidad a partir de los Eigen-CAM³², teniendo en cuenta su relativa independencia a la probabilidad de salida y su posible contribución como regularizador de los nódulos candidatos. Este método consiste en utilizar los componentes principales de las representaciones aprendidas en las capas convolucionales. Esta estrategia de explicabilidad es independiente de la probabilidad de salida de una imagen en particular, tomando como base de representación, el conjunto de activaciones convolucionales en una capa específica, especialmente donde los componentes de salida son preservados.

En este sentido, el conjunto de activaciones convolucionales son tomadas como entrada para formar una covarianza, y cuya descomposición espectral, permite al mejor eje de varianza de las características convolucionales, sin la necesidad de modificar la arquitectura convolucional o incluso usar la retropropagación. Teniendo una imagen de entrada representada por $\mathbf{I} \in R^{K \times M}$, se utilizan las diferentes activaciones representada $\mathbf{F} \in R^{O \times P}$ en la capa de interés, en donde el total de activaciones tendrían la siguiente representación $\mathbf{W} \in R^{O \times P \times C}$ primero se cambia la forma de \mathbf{W} para que sea una matriz bidimensional teniendo la siguiente representación $\mathbf{A} \in R^{B \times C}$, donde cada fila representa una ubicación en la imagen y cada columna representa la activación de un filtro convolucional. Luego, se centra la matriz \mathbf{A} restando la media de cada columna. Después obtener esta matriz bidimensional \mathbf{A} se le hace la descomposición en valores singulares (por sus siglas en inglés: SVD). $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$, en donde \mathbf{U} es una matriz ortogonal $\mathbf{U} \in R^{B \times B}$, Σ una matriz diago-

³² MUHAMMAD, Mohammed Bany y YEASIN, Mohammed. "Eigen-cam: Class activation map using principal components". En: *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2020, págs. 1-7.

nal de tamaño $\Sigma(B \times C) \in R^{B \times C}$, V una matriz ortogonal de tamaño $V \in R^{C \times C}$, Obteniendo el mapa de clase mediante componente principales viene dada por la proyección de A sobre el primer vector propio; $Eigen-CAM = AV_1$ donde el primer vector propio de V tiene una mejor representación de las características principales como se observa en la figura 8.

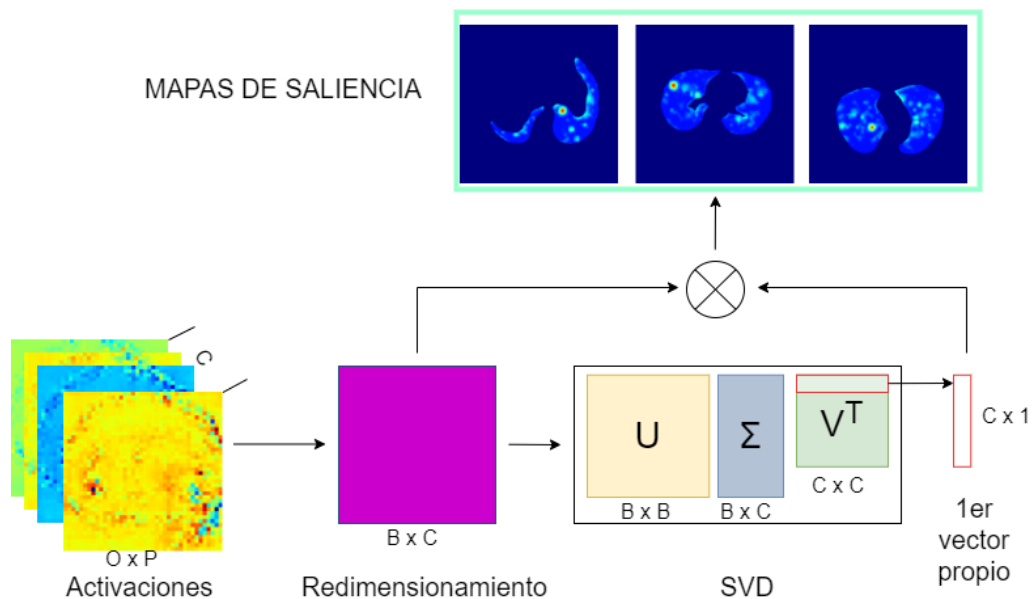


Figura 8. Mapas saliencia por medio de componente principales.

En la Figura 9 se muestran los mapas obtenidos desde las capas convolucionales primarias. Es importante notar, que el método se aplica en estas capas para tener una menor dependencia de la supervisión y lograr hacer el análisis de variabilidad sobre representaciones primarias de las imágenes.

4.4. Generación de nódulos candidatos regularizados por atención

La fase de detección, también denominada generación de cuadros delimitadores, se utiliza al final de la arquitectura con el fin de generar las detecciones de los nódulos pulmonares. Esta generación de cuadros delimitadores aplica X cuadros de anclaje

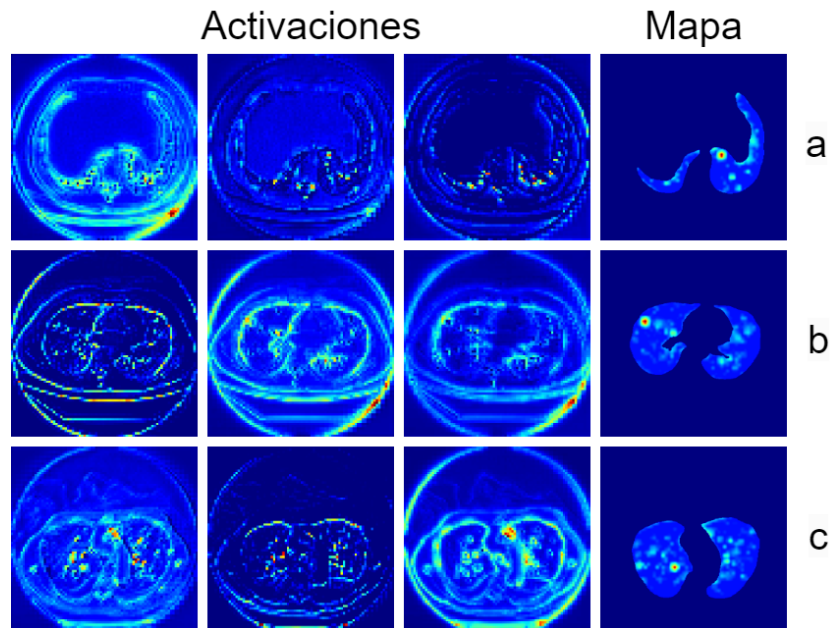


Figura 9. Activaciones de los mapas de componente principales y sus mapas de saliencia en donde a) hay algunas discrepancias en las regiones detectadas entre las activaciones y los mapas, b) las activaciones concuerdan mayormente con lo obtenido en los mapas, c) casi no concuerdan las activaciones con los mapas.

denotados como x_1, x_2, \dots, x_n los cuales son establecidos desde la proyección inicial a la arquitectura de la red para que sean adecuados para cualquier tamaño de nódulos, sobre la salida del mapa de características de la capa anterior y luego genera un vector con la probabilidad de clase del objeto. Además, cada cuadro delimitador contará con una probabilidad de contener un nódulo, así como también sus respectivas coordenadas.

En este trabajo en particular se utiliza únicamente la salida final como representación para detectar nódulos, así como también una reducción de falsos positivos complementados por los mapas de saliencia como se puede observar en la Figura 6 la salida de representación, que brinda una mayor sensibilidad está ubicada en la etapa 3. Para conocer la importancia que puede tener los cuadros de anclaje a la

función de pérdida se va a describir a continuación como: $b = \sigma(t) + c$ y $a = pe^t$ donde b es la posición donde se encuentra el nódulo descrito por sus coordenada $b = [b_x, b_y]$, a es el ancho y alto de la caja delimitadora prevista descrito como $a = [a_h, a_w]$, t como las salidas de la red tras el entrenamiento descrito como $t = [t_x, t_y, t_h, t_w]$, c como las coordenadas de la esquina superior izquierda de la cuadrícula de la grilla donde se realiza la predicción descrito como $c = [c_x, c_y]$ y p como los cuadros de anclaje definidos al principio de la red descritos como $p = [p_h, p_w]$.

Alternativamente, como una contribución de este trabajo se propuso una detección adicional de nódulos pulmonares utilizando los eigen-mapas de explicabilidad. Estos mapas son fusionados con los nódulos detectados por la representación aprendida y así permitiendo mitigar los falsos positivos en la representación supervisada.

Primero, para este método alternativo se requiere una segmentación de los pulmones, la cual se realizó mediante la U-Net (R231)³³. Esta etapa de segmentación tiene como objetivo aplicar un filtro a los mapas de saliencia generados, a los cuales se les dibujan cuadros delimitadores en los valores de intensidad que superan el umbral del 96 %. Este umbral corresponde a la región de mayor importancia para la red, y donde intuitivamente se podría sospechar que se encuentra un nódulo pulmonar. Estos cuadros delimitadores se comparan con la etiqueta real para determinar si se pudo detectar el nódulo pulmonar. Una ilustración de este proceso se puede ver en la Figura 10.

Además, se decidió que los cuadros delimitadoras debían ser dinámicos. Es decir, al dibujar la primera caja delimitadora, esta se compara con la etiqueta real y se esta-

³³ HOFMANNINGER, Johannes, *et al.* "Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem". En: *European Radiology Experimental* 4.1 (2020), págs. 1-13.

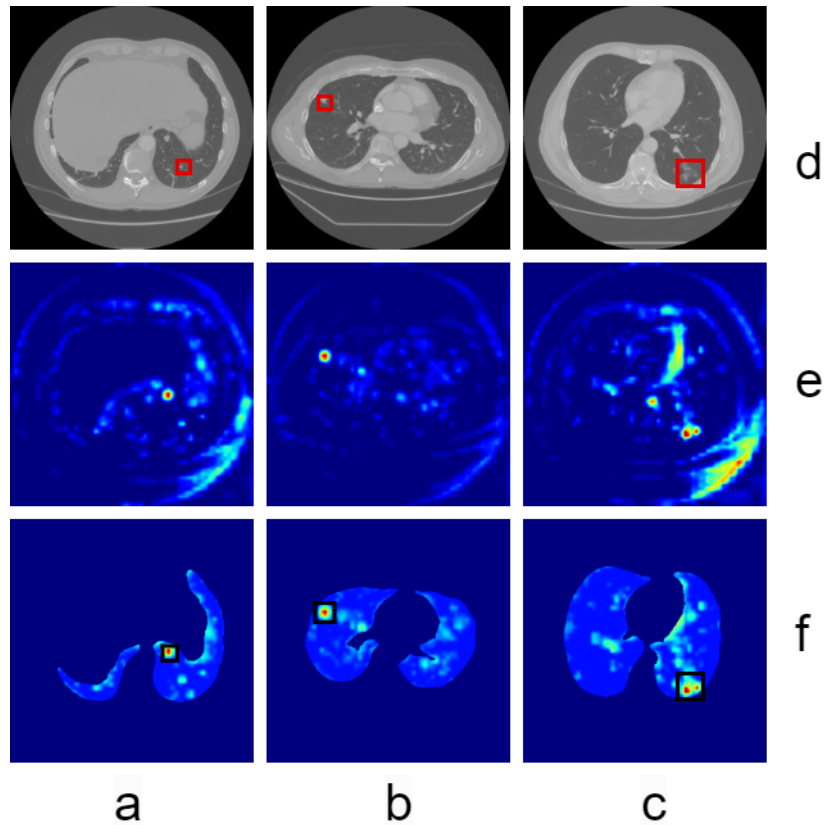


Figura 10. Activaciones de los mapas de componente principales y sus mapas de saliencia en donde a) La región que detecta el mapa no es un nódulo b) el Mapa si localiza el nódulo c) La segmentación quita los borde, que ayuda a mejorar la localización, d) Corte tomografía computarizada, e) Mapas de saliencia, f) Mapa filtrado o segmentado.

blece el valor mínimo de intensidad en el mapa de saliencia. Luego, para generar la próxima caja delimitadora, se utiliza el valor de intensidad que es igual o superior al 96 % del mapa de saliencia. Estos mapas de saliencia tienen la capacidad de disminuir la ocurrencia de resultados positivos incorrectos en las predicciones del modelo.

Para su implementación, se procedió a evaluar las predicciones generadas por la arquitectura YOLO propuesta, y se llevó a cabo una comparación con los mapas Eigen generados. Este enfoque involucró la obtención de todas las predicciones realizadas por la red propuesta, incluyendo aquellas clasificadas como falsos posi-

vos. A continuación, se analizaron estas predicciones en conjunto con las regiones destacadas por los mapas de saliencia, seleccionando aquellas áreas que presentaban una mayor intensidad, lo cual indicaba una mayor probabilidad de presencia de nódulos.

En caso de que la intensidad en el mapa de saliencia no superara un umbral previamente establecido, se descartaba la predicción. En caso contrario, se realizaba una comparación con la predicción original de la red, con el fin de determinar si se correspondía con lo esperado. De esta forma, se logró reducir los falsos positivos obtenidos por la red.

5. CONFIGURACIÓN EXPERIMENTAL

El enfoque propuesto se evaluó a partir de un conjunto de datos públicos LIDC-IDRI (Lung Image Database Consortium Image Collection)³⁴. Este conjunto de datos ha sido ampliamente usado en la literatura, teniendo como principal ventaja una gran cantidad de información de los nódulos, relacionado con su grado de severidad, caracterización morfológica y asociación patológica.

Particularmente el LIDC-IDRI cada estudio CT tiene una dimensión espacial de 512×512 y en promedio cada estudio cuenta 200 cortes. Las observaciones nodulares se realizaron por un total de cuatro radiólogos experimentados, que reportan más de 5 años de experiencia. A demás siguiendo un criterio LUNA16³⁵, los nódulos seleccionados para entrenar o validar el estudio era aquellos que contaban con observaciones positivas por al menos 3 radiólogos. Igualmente, el criterio LUNA solo considera nódulos positivos a aquellos que reportan un diámetro superior a 3 mm. En este caso, de la base total de nódulos se tomaron 1187 nódulos. Para esta toma de datos se utilizaron una variedad de fabricantes y modelos de escáneres como lo son GE Medical Systems LightSpeed, Philips Brilliance, Siemens Definition, Emotion, Sensation y Toshiba Aquilion. El tamaño de píxel en el plano osciló entre 0,461 y 0,977 mm (media: 0,688 mm). En estudio hubo 1010 pacientes diferentes entre hombres y mujeres, la intención era incluir solo un escaneo de cualquier paciente para que los escaneos en la base de datos no estuvieran correlacionados. En la siguiente gráfica se puede ver unas imágenes TC proporcionadas por el conjunto

³⁴ ARMATO III, Samuel G, *et al.* "The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans". En: *Medical physics* 38.2 (2011), págs. 915-931.

³⁵ <https://luna16.grand-challenge.org/Data/>.

de datos LIDC-IDRI 11.

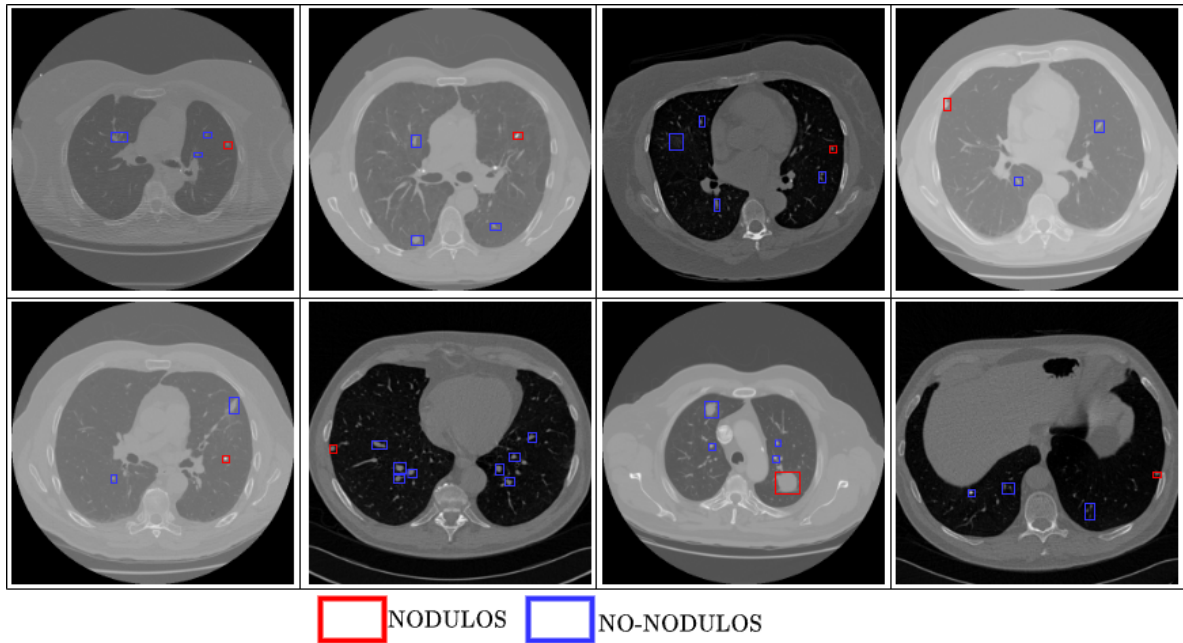


Figura 11. Imágenes TC proporcionados por el conjunto de datos LIDC-IDRI donde los cuadros rojos son los nódulos seleccionados por 3 o más radiólogos y los azules son aquellos que fueron seleccionados por menos de 3 radiólogos o falsos positivos. Como se evidencia, los cortes tienen una alta variabilidad, diferentes contrastes y los nódulos pueden ser altamente variables.

Para la distribución de nuestros datos se decidió utilizar el 70 % para el entrenamiento, el 10 % para la validación y el 20 % para la prueba.

5.1. Configuración del modelo

El modelo implementado se basó en la versión YOLOv5s, utilizando los hiper-parámetros ya predefinidos por la red, tales como 212 capas, 8 módulos C3, 1 modulo SPPF. Además esta arquitectura tiene un total de 7 millones de parámetros entrenables. En cuanto al entrenamiento se siguió una optimización con 300 épocas y un batch de 32 cortes. Para el entrenamiento se utilizó un optimizador descenso del gradiente estocástico (por sus siglas en inglés: SGD) con un learning rate de 0.01, se usaron

pesos inicializados aleatoriamente por la red con un decaimiento de peso (por su nombre en inglés: Weight Decay) de 0.0005, un momentum de 0.937, un umbral de IOU de 0.3 y un umbral de sensibilidad variable.

Por otra parte, se utilizó solo una grilla de entrada de 16×16 por ser la que genera el menor número de predicciones en cada imagen además de ser la más precisa, pero como dejaba un pequeño número de nódulos sin detectar, se decidió aumentar un poco este número de predicciones aumentando el número de cuadros de anclaje a 9 al ser el número más óptimo para realizar la detección de nódulos. Estos cuadros de anclaje tienen las siguientes medidas [10×13 , 16×30 y 33×23] generando 3 cuadros de cada una.

En cuanto a la definición de mapas de saliencia, se seleccionó la etapa 1 del bloque c3 que cuenta con 128 filtros y una matriz de convolución 3×3 . Luego se hizo un redimensionamiento a las activaciones y una descomposición siguiendo los valores singulares, se obtuvieron un mapa por cada corte de entrada. Para hallar el umbral de intensidad se varió desde un valor mínimo 0% hasta el valor máximo 100% obteniendo el valor más óptimo en 96%.

6. EVALUACIÓN Y RESULTADOS

A continuación, se presentarán la evaluación y resultados de la metodología propuesta, haciendo un análisis de componentes y evidenciando la capacidad de localizar nódulos sobre cortes de TC. Esta propuesta fue valida completamente sobre el conjunto de datos LIDC-IDRI. Primero se presentara el estudio exhaustivo que se hizo para escoger cual iba hacer la arquitectura base, después cuales características se le modificaron para que fuera óptima para la detección de nódulos pulmonares como lo es la etapa de detección más idónea para realizar la localización, el tamaño y numero adecuado de cuadros de anclaje para la red, posteriormente se presentara el aporte que brindan los mapas de saliencia a la red y finalizar con la comparación respecto al estado del arte y algunos ejemplos de los mapas generados. Los resultados obtenidos de presentaran mediante gráficas y tablas siguiendo la métrica de rendimiento competitivo (CPM) el cual está definido como la sensibilidad media de siete tasas de falsos positivos por imagen y la curva FROC la cual es muy utilizada para proyectos enfocados en la localización.

Estos resultados son muy interesantes, porque se encuentran muchas variaciones de la arquitectura YOLO, además de cómo se puede comparar con grandes métodos de localización de nódulos en el estado del arte y el gran aporte que pueden brindan los mapas de saliencia a una red para poder entenderla mejor y poder obtener mejores resultados. En primer lugar, comparamos las diferentes versiones de la YOLO, como la YOLOv3, YOLOv5 y YOLOv7, así como algunas de sus sub familias como la YOLOv3-SPP, YOLOv5s, YOLOv5l, YOLOv5x para corroborar cuál es la versión más óptima para realizar la localización de nódulos. Los resultados se pueden visualizar en la curva FROC 12 la cual compara la sensibilidad respecto a la tasa de los distintos falsos positivos por imagen.

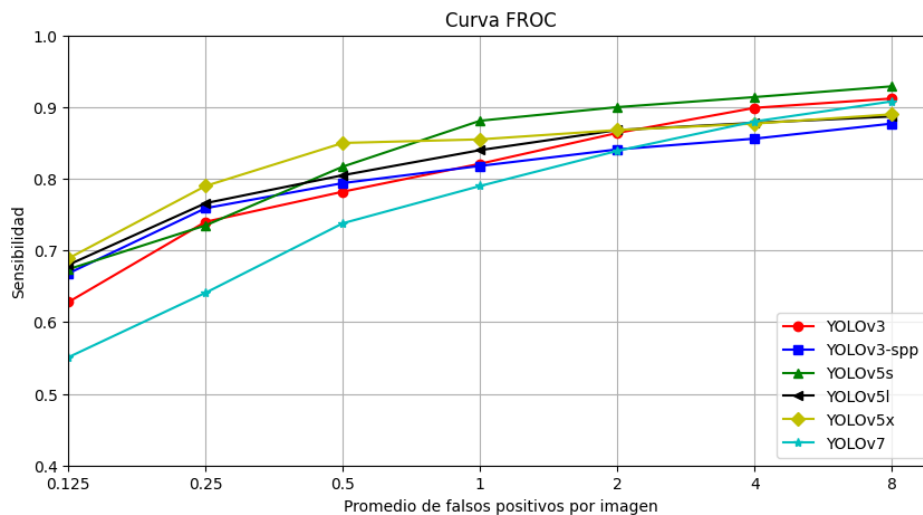


Figura 12. Curva FROC de los distintos modelos de la YOLO

De acuerdo a la figura 12, la arquitectura YOLOv5s reporta un mejor comportamiento para la caracterización de nódulos. Esto se debe a su extractor de características único de la YOLOv5 que preserva mejor la información espacial además de sus distintos canales de procesamiento al momento de hacer la localización y que no necesita ser muy profunda en comparación con las otras versiones de la YOLOv5. A continuación, para conocer la contribución de las diferentes etapas de detección de la red, se corroboró el funcionamiento de cada una de ellas sin utilizar las otras etapas y con todas las etapas juntas. Una descripción de las etapas se ilustra en el marco general del método propuesto (ver figura 6). En el gráfico de barras 13 se puede ver la comparación de cada etapa diferente y de su eficacia para detectar nódulos.

En la última etapa de detección, se obtienen mejores resultados en comparación con las otras 2 etapas, incluso mejora en comparación con la YOLOv5s con las 3 etapas juntas, esto puede deberse a que al estar al final de la red se aprenden características importantes de los nódulos que las otras 2 no pueden, además al tener un tamaño de grilla menor se generan menos detecciones innecesarias que

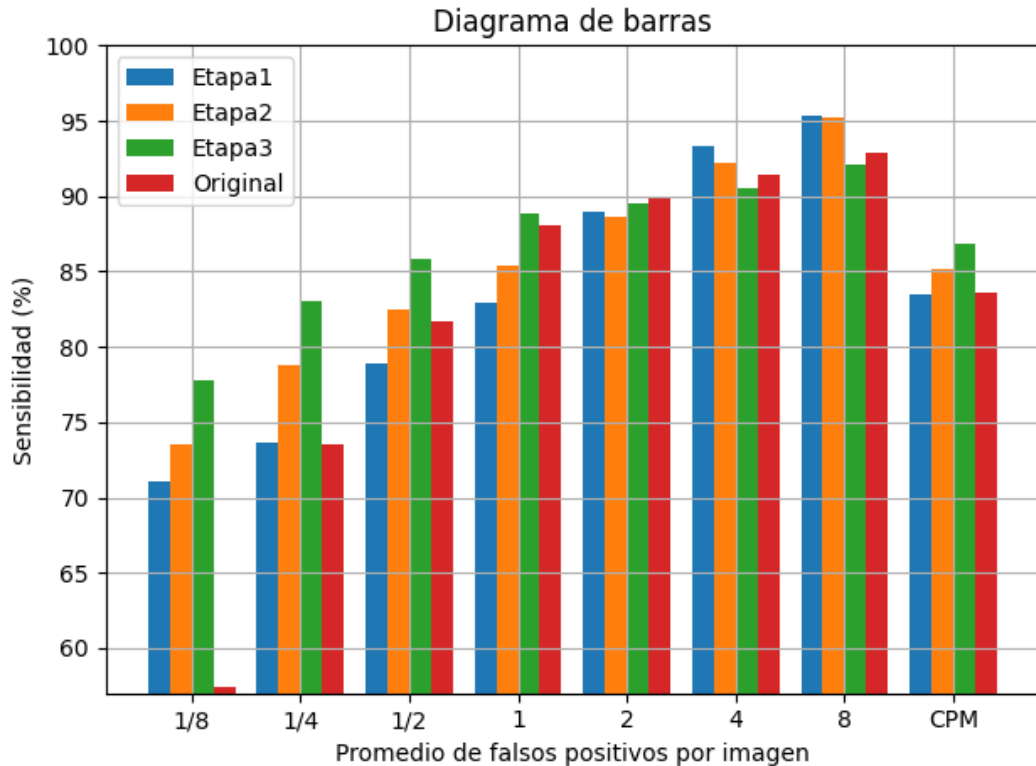


Figura 13. Modelo de barras comparando cada una de las etapas de la red y la red original.

las otras 2 pueden estar generando o que generan las 3 etapas juntas.

A través de la función de pérdida utilizada por la YOLOv5s para elegir el punto central y el cuadro delimitador adecuado para los nódulos, se observó la importancia de los cuadros de anclaje, por lo tanto, el tamaño de los cuadros de anclaje se varió entre mediano, pequeño y muy pequeño. Como la YOLOv5s en el entrenamiento compara el tamaño de los cuadros en las etiquetas respecto a los cuadros de anclaje propuesto no permitirá que sean muy grandes. En la tabla 1 se utiliza para comprobar cuál de estos diferentes tipos de tamaños es el más adecuado para la detección de nódulos.

Como era de esperar, el tamaño adecuado fue el de los cuadros de anclaje con un tamaño pequeño, ya que hay nódulos que pueden ser tanto muy pequeños, así como nódulos de tamaño mediano, lo que permite que los cuadros de anclaje se

Tabla 1. YOLOv5s con diferentes tamaños de cuadros de anclaje, donde los cuadros medianos tienen un tamaño mayor a 33×23 píxeles, los pequeños de $[10 \times 13, 16 \times 30$ y $33 \times 23]$ como se especificó en la configuración del modelo y los muy pequeños menores a 10×13 píxeles.

TAMAÑO	Sensibilidad respecto a los de falsos positivos por imagen(%)							CPM
	1/8	1/4	1/2	1	2	4	8	
MEDIANOS	76.11	79.61	84.81	87.69	90.42	92.22	94.30	86.45
PEQUEÑOS	75.81	79.60	84.75	89.62	92.25	93.50	95.33	87.27
MUY PEQUEÑOS	74.23	77.52	83.40	89.55	92.30	93.92	95.90	86.69

ajusten mejor a la función de pérdida, aunque no hubo cambios significativos entre estos tipos de tamaño porque se eligieron por ser relativamente adecuados para los nódulos.

Seguidamente se probó el número adecuado de cuadros de anclaje para mejorar la detección de nódulos. En la tabla 2 podemos comparar el rendimiento de la red en su última etapa de detección con diferentes números de cuadros de anclaje.

Tabla 2. YOLOv5s con diferentes números de cuadros de anclaje.

CUADROS DE ANCLAJE	Sensibilidad respecto a los de falsos positivos por imagen(%)							CPM
	1/8	1/4	1/2	1	2	4	8	
3	79.61	83.46	87.06	88.20	89.70	90.60	91.15	87.11
6	80.38	83.81	85.80	86.80	87.80	89.10	89.84	86.22
9	79.03	84.08	87.08	89.01	90.87	92.19	93.64	87.99
12	81.30	83.96	87.13	88.73	89.10	89.10	89.50	86.97

Aunque los resultados variando el número de cuadros de anclaje no es muy significativo, se obtienen mejores resultados con 9 de ellos porque si se aumenta este número los resultados van disminuyendo. Esto es debido a que al tener muy pocas cuadros de anclaje, algunos nódulos pueden no ser detectados en la grilla, y al tener muchas, la grilla puede no elegir la más óptima para el nódulo.

De manera alternativa y además complementaria, en este trabajo se exploró el uso de los mapas de saliencia con eigen-mapas para la detección de nódulos. En este caso, al observar en los mapas de saliencia que la mayor intensidad resaltaba el nódulo, se calcularon mapas de saliencia para cada estado, en los que se obtiene

un porcentaje de casi el 75% CPM como se puede visualizar en la tabla 3. Se puede observar que, aunque no supera el CPM de la arquitectura propuesta aun así nos da una buena representación de los nódulos. para posibles trabajos futuros se podría dar más relevancia a estos mapas de saliencia, pero teniendo en cuenta que computacionalmente no es posible obtener los porcentajes de $\frac{1}{8}$ $\frac{1}{4}$ $\frac{1}{2}$ en la métrica, ya que al iterar sobre ella tomando el umbral del 96%, el porcentaje en todos los datos de entrenamiento oscila alrededor de 1 falso positivo por imagen.

Tabla 3. Sensibilidad obtenida de los mapas de saliencia en diferentes etapas.

Etapas	Sensibilidad respecto a los de falsos positivos por imagen				CPM
	1	2	4	8	
Etapa3	0.502	0.672	0.807	0.895	0.719
Etapa2	0.519	0.704	0.835	0.916	0.744
Etapa1	0.510	0.713	0.840	0.916	0.745

A demás se pueden utilizar estos mapas de saliencia para disminuir el número de falsos positivos esto se puede hacer al comparar los mapas de saliencia generados y las predicciones de la red, se hizo posible reducir el porcentaje de falsos positivos como se puede ver en la Figura 14.

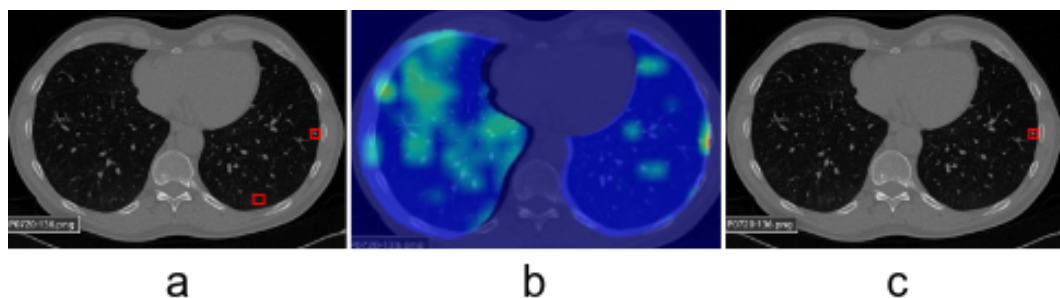


Figura 14. a) Representa la predicción propuesta por la arquitectura. b) nos muestra las regiones de interés propuesta por el mapa de saliencia. c) corresponde a la etiqueta real del nódulo.

Gracias a los mapas de saliencia propuestos, podemos notar algunas regiones donde no estaría el nódulo pero que la red lo predice erróneamente, de esta manera

comparándolos podemos hacer una reducción de falsos positivos. De este modo la red mejora en su detección, aunque sea en un pequeño porcentaje como se puede ver en la tabla 4. Aunque no mejora su sensibilidad cuando se generan pocos falsos positivos se puede ver su aporte cuantos más falsos positivos existan.

Tabla 4. Modelo propuesto con reducción de falsos positivos. Mejora en el rendimiento al complementar los mapas de saliencia con un mayor numero de falsos positivos.

Modelos	CPM
Red propuesta sin mapas de saliencia	87.99
Red propuesta con mapas de saliencia	88.01

Posteriormente se intentó hacer una comparación con otros trabajos del estado del arte. Una primera dificultad encontrada es la limitación de realizar una validación con los mismos datos en las particiones. A pesar que los trabajos son reportados bajo el mismo dataset las particiones de estos datos pueden variar con cada uno de los autores. Los autores además han reportado puntajes globales en discriminar entre los tamaños de los nódulos, un punto vital en el análisis de estos. Algunos trabajos del estado del arte han reportado diferentes resultados como por ejemplo Atrous Unet ²² (92 % para nódulos menores a 9 mm y 93 % para nódulos mayores a 10 mm, usando un test del 20 % de las muestras), Faster ¹⁹ (77.5 %, usando una validación cruzada de 10), Unet ²¹ (92.5 % usando una validación cruzada de 10). En este sentido nuestro trabajo obtuvo un puntaje general de CPM de 88.01 %, usando un total de muestras de test del 20 % de los datos. En nuestro caso, el resultado de CPM para nódulos menores de 8 mm fue de 87.30 %, mientras que para nódulos mayores de 8 mm fue de 89.50 %. Cabe destacar en estos resultados, que se obtiene un resultado similar para nódulos con diferentes comportamientos, y que además su puntaje global es competitivo con el estado del arte.

De manera complementaria, los mapas generados pueden ser una alternativa para el análisis observacional de los nódulos. En la arquitectura propuesta generamos mapas de saliencia utilizando la técnica Eigen-CAM, observando visualmente en la Figura 15, estos mapas son capaces de resaltar las regiones que más contribuyen a la detección de nódulos pulmonares, incluso en muchos mapas de saliencia la región de mayor intensidad representa un nódulo pulmonar en el corte seleccionado. Además, cabe destacar que estos mapas de saliencia contienen información muy importante en la detección de nódulos pulmonares, que ofrecería información extra a los radiólogos para su análisis. Como se puede observar en 15 [a], se observa un buen mapa de saliencia, con un valor de intensidad mayor de 175 el cual resalta el nódulo, mientras que en [b], tenemos otro buen mapa de saliencia con un valor de intensidad mayor de 175 pero que no resalta el nódulo si no por el contrario nos muestra un falso positivo, en [c] tenemos un mapa de saliencia ruidoso, pero consigue detectar el nódulo pulmonar, al realizar muchas iteraciones en la creación de los cuadros delimitadores, [d], el mapa de saliencia es ruidoso pero contrario al anterior no es capaz de detectar el nódulo pulmonar.

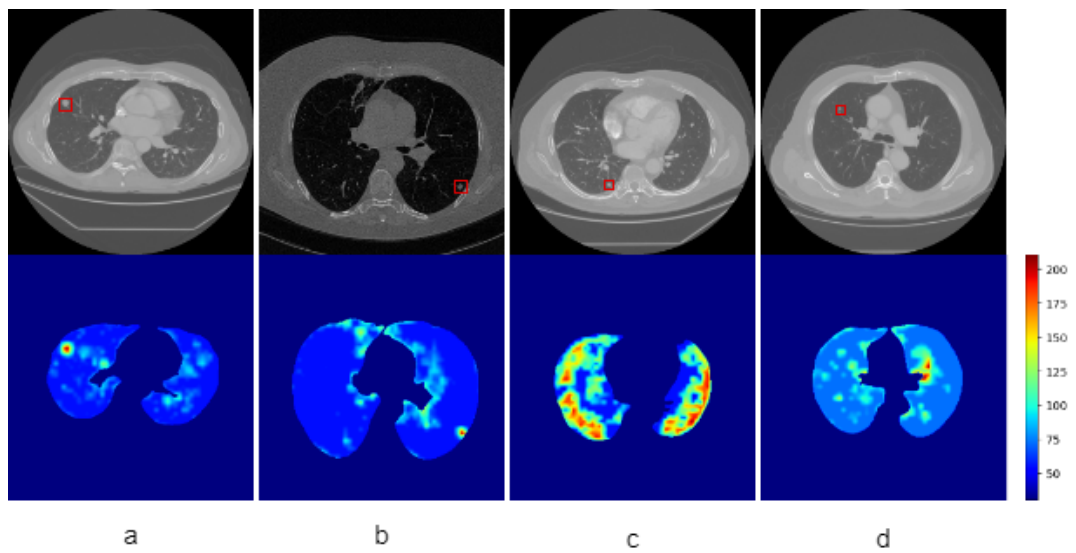


Figura 15. Se pueden observar distintos mapas de saliencia obtenidos por la red. Donde a) un mapa poco ruidoso que resalta el nódulo, b) un mapa poco ruidoso, pero con muchos falsos positivos, c) un mapa muy ruidoso que resalta el nódulo, d) un mapa muy ruidoso y con muchos falsos positivos.

7. CONCLUSIONES Y PERSPECTIVAS

En este trabajo se implementó una arquitectura convolucional basada en la Yolov5s, para la detección de nódulos en cortes de TC. Esta arquitectura tiene una amplia capacidad para preservar la espacialidad de las regiones de interés. Además, de manera complementaria se calcularon mapas de saliencia que armonizan la detección supervisada y ayudan a la reducción de falsos positivos. El método implementado fue validado y comparado con diferentes arquitecturas del estado del arte en localización de nódulos. Del mismo modo, su ajuste y validación fue llevado a cabo con un conjunto de datos publico. Los resultados muestran un comportamiento de 88.01 %.

En primer lugar, se realizó un estudio exhaustivo y se visualizaron mapas Eigen-CAM para cada etapa de detección de la red, gracias a esto se observó que la última etapa proporciona mejor información discriminatoria que las etapas anteriores, por lo que esta última etapa es la que más contribuye a la localización de nódulos pulmonares, por lo que se aumentó el número de cuadros de anclaje para esta etapa y se eliminaron las etapas anteriores de detección. En segundo lugar, se cambió el tamaño de los cuadros de anclaje por otros más adecuadas para la detección de nódulos, optimizando así la función de pérdida. Por otra parte, los mapas de saliencia se utilizaron como alternativa de localización y ayudaron a disminuir el porcentaje de falsos positivos que pueden surgir en la red.

En este trabajo podemos concluir que es competitivo en el estado del arte pero que además puede brindar información especial que otras redes no brindan mediante el uso del Eigen-CAM lo que puede ser de gran importancia para la detección de nódulos. Pero aún tiene margen para mejorar su sensibilidad y precisión estudian-

do diferentes alternativas en mayor profundidad. Para trabajos futuros esperamos poder utilizar distintos pasos de reducción de falsos positivos, teniendo en cuenta las diferentes características del nódulo y su área circundante como lo puede ser utilizando los mapas de saliencia y así implementar una red neuronal residual (por su nombre en inglés: Residual neural network) en donde posiblemente ayudaría a reducir muchos falsos positivos, o mejorando algunos bloques de la YOLOv5s. En la parte de la concatenaciones implementar mecanismos de atención, así como ampliando el alcance del estudio aumentando el número de datos para cumplir con los requisitos clínicos y poder ser llevado a un área de aplicación.

BIBLIOGRAFÍA

AGNES, S Akila; ANITHA, J y SOLOMON, A Arun. “Two-stage lung nodule detection framework using enhanced UNet and convolutional LSTM networks in CT images”. En: *Computers in Biology and Medicine* 149 (2022), pág. 106059 (vid. págs. 25, 49).

ALBERT, Ross H y RUSSELL, John J. “Evaluation of the solitary pulmonary nodule”. En: *American family physician* 80.8 (2009), págs. 827-831 (vid. pág. 29).

ARMATO III, Samuel G, *et al.* “The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans”. En: *Medical physics* 38.2 (2011), págs. 915-931 (vid. pág. 41).

BUSH, Isabel. “Lung nodule detection and classification”. En: *Rep. Stanf. Comput. Sci* 20 (2016), págs. 196-209 (vid. pág. 26).

CAO, Haichao , *et al.* “A two-stage convolutional neural networks for lung nodule detection”. En: *IEEE journal of biomedical and health informatics* 24.7 (2020), págs. 2006-2015 (vid. págs. 24, 49).

DEL CIELLO, Annemilia, *et al.* “Missed lung cancer: when, where, and why?” En: *Diagnostic and interventional radiology* 23.2 (2017), pág. 118 (vid. págs. 13, 28).

DONG, Xiao, *et al.* “An improved YOLOv5 network for lung nodule detection”. En: *2021 International Conference on Electronic Information Engineering and Computer Science (EIECS)*. IEEE. 2021, págs. 733-736 (vid. pág. 26).

DUMA, Narjust; SANTANA-DAVILA, Rafael y MOLINA, Julian R. “Non–small cell lung cancer: epidemiology, screening, diagnosis, and treatment”. En: *Mayo Clinic Proceedings*. Vol. 94. 8. Elsevier. 2019, págs. 1623-1640 (vid. págs. 12, 28).

FAN, Jianqing; MA, Cong y ZHONG, Yiqiao. “A selective overview of deep learning”. En: *Statistical science: a review journal of the Institute of Mathematical Statistics* 36.2 (2021), pág. 264 (vid. pág. 17).

FENG, Xinyang, *et al.* “Discriminative localization in CNNs for weakly-supervised segmentation of pulmonary nodules”. En: *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III* 20. Springer. 2017, págs. 568-576 (vid. pág. 27).

GASCOIGNE-BURNS, James y KATSIGIANNIS, Stamos. “A Localisation Study of Deep Learning Models for Chest X-ray Image Classification”. En: *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE. 2022, págs. 01-05 (vid. pág. 26).

GONG, Li, *et al.* “Automated pulmonary nodule detection in CT images using 3D deep squeeze-and-excitation networks”. En: *International journal of computer assisted radiology and surgery* 14 (2019), págs. 1969-1979 (vid. pág. 25).

HAN, Hao, *et al.* “Fast and adaptive detection of pulmonary nodules in thoracic CT images using a hierarchical vector quantization scheme”. En: *IEEE journal of biomedical and health informatics* 19.2 (2014), págs. 648-659 (vid. pág. 13).

HOFMANNINGER, Johannes, *et al.* “Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem”. En: *European Radiology Experimental* 4.1 (2020), págs. 1-13 (vid. pág. 38).

HUANG, Rachel; PEDOEEM, Jonathan y CHEN, Cuixian. “YOLO-LITE: a real-time object detection algorithm optimized for non-GPU computers”. En: *2018 IEEE international conference on big data (big data)*. IEEE. 2018, págs. 2503-2510 (vid. pág. 20).

HUANG, Xia, *et al.* “Fast and fully-automated detection and segmentation of pulmonary nodules in thoracic CT scans using deep convolutional neural networks”. En: *Computerized Medical Imaging and Graphics* 74 (2019), págs. 25-36 (vid. pág. 24).

LIU, Chenyang, *et al.* “Automatic detection of pulmonary nodules on CT images with YOLOv3: development and evaluation using simulated and patient data”. En: *Quantitative Imaging in Medicine and Surgery* 10.10 (2020), pág. 1917 (vid. págs. 13, 26).

LIU, Kai, *et al.* “Evaluating a fully automated pulmonary nodule detection approach and its impact on radiologist performance”. En: *Radiology: Artificial Intelligence* 1.3 (2019), e180084 (vid. pág. 13).

LIU, Kehong. “STBi-YOLO: A Real-Time Object Detection Method for Lung Nodule Recognition”. En: *IEEE Access* 10 (2022), págs. 75385-75394 (vid. pág. 25).

MEI, Sen; JIANG, HuiQin y MA, Ling. “YOLO-lung: a practical detector based on improved YOLOv4 for pulmonary nodule detection”. En: *2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE. 2021, págs. 1-6 (vid. pág. 26).

MONKAM, Patrice, *et al.* "Detection and Classification of Pulmonary Nodules Using Convolutional Neural Networks: A Survey". En: *IEEE Access* 7 (2019), págs. 78075-78091. DOI: 10.1109/ACCESS.2019.2920980 (vid. pág. 12).

MUHAMMAD, Mohammed Bany y YEASIN, Mohammed. "Eigen-cam: Class activation map using principal components". En: *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2020, págs. 1-7 (vid. pág. 35).

NGUYEN, Chi Cuong, *et al.* "Pulmonary nodule detection based on faster R-CNN with adaptive anchor box". En: *IEEE Access* 9 (2021), págs. 154740-154751 (vid. pág. 24).

PENG, Haixin; SUN, Huacong y GUO, Yanfei. "3D multi-scale deep convolutional neural networks for pulmonary nodule detection". En: *Plos one* 16.1 (2021), e0244406 (vid. pág. 25).

RUBIN, Geoffrey D. "Lung nodule and cancer detection in CT screening". En: *Journal of thoracic imaging* 30.2 (2015), pág. 130 (vid. pág. 13).

SELVARAJU, Ramprasaath R, *et al.* "Grad-CAM: Why did you say that?" En: *arXiv preprint arXiv:1611.07450* (2016) (vid. págs. 22, 23).

SHI, Lukui; MA, Hongqi y ZHANG, Jun. "Automatic detection of pulmonary nodules in CT images based on 3D Res-1 network". En: *The Visual Computer* 37 (2021), págs. 1343-1356 (vid. pág. 14).

SUNG, Hyuna, *et al.* "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries". En: *CA: a cancer journal for clinicians* 71.3 (2021), págs. 209-249 (vid. págs. 12, 28).

WU, Wentong, *et al.* “Application of local fully Convolutional Neural Network combined with YOLO v5 algorithm in small target detection of remote sensing image”. En: *PloS one* 16.10 (2021), e0259283 (vid. pág. 20).

XIE, Hongtao, *et al.* “Automated pulmonary nodule detection in CT images using deep convolutional neural networks”. En: *Pattern Recognition* 85 (2019), págs. 109-119 (vid. págs. 24, 49).

ZHANG, Guobin, *et al.* “Automatic nodule detection for lung cancer in CT images: A review”. En: *Computers in biology and medicine* 103 (2018), págs. 287-300 (vid. págs. 17, 29).

ZHOU, Bolei, *et al.* “Learning deep features for discriminative localization”. En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, págs. 2921-2929 (vid. págs. 21-23).