

Pontificia Universidad Javeriana Cali
Facultad de Ingeniería.
Ingeniería de Sistemas y Computación.
Trabajo de Grado.

Detección automática de masas en mamografías digitales mediante técnicas de inteligencia artificial

Fabian Antoyne Garcia Gallego

Director: MSc. Felipe Palta

13/01/2026



Santiago de Cali, 13/01/2026.

Señores

Pontificia Universidad Javeriana Cali.

Dr. Gerardo Mauricio Sarria

Director Carrera de Ingeniería de Sistemas y Computación.

Cali.

Cordial Saludo.

Por medio de la presente me permito informarle que el estudiante de Ingeniería de Sistemas y Computación Fabian Antoyne Garcia Gallego (cod: 8956118) trabaja bajo mi dirección en el proyecto de grado titulado “Detección automática de masas en mamografías digitales mediante técnicas de inteligencia artificial”.

Atentamente,

Felipe Palta

MSc. Felipe Palta

Santiago de Cali, 13/01/2026.

Señores

Pontificia Universidad Javeriana Cali.

Dr. Gerardo Mauricio Sarria

Director Carrera de Ingeniería de Sistemas y Computación.

Cali.

Cordial Saludo.

Me permito presentar a su consideración el trabajo de grado titulado “Detección automática de masas en mamografías digitales mediante técnicas de inteligencia artificial” con el fin de cumplir con los requisitos exigidos por la Universidad para llevar a cabo el proyecto de grado y posteriormente optar al título de Ingeniero de Sistemas y Computación.

Al firmar aquí, doy fe que entiendo y conozco las directrices para la presentación de trabajos de grado de la Facultad de Ingeniería aprobadas el 26 de Noviembre de 2009, donde se establecen los plazos y normas para el desarrollo del anteproyecto y del trabajo de grado.

Atentamente,



Fabian Antoyne Garcia Gallego

Código: 8956118

Abstract

Early detection of breast cancer through mammography is a problem of high clinical impact; however, manual localization of masses can be laborious due to low contrast, tissue overlap, and high morphological variability of lesions. This work presents an automatic mass detection system in digitized mammograms using the one-stage YOLO family of detectors. The mass subset of the CBIS-DDSM dataset was used to build a preprocessing pipeline that (i) loads mammograms in DICOM format, (ii) uses ROI masks to derive consistent bounding boxes, and (iii) generates annotations in YOLO format and training, validation, and test partitions.

A pre-trained YOLOv8m model was selected as the base model and fine-tuned via transfer learning to a single class (mass) with `imgsz=640`. On the validation set, the model achieved $P = 0,586$, $R = 0,544$, $\text{mAP}@0,5 = 0,533$, and $\text{mAP}@0,5 : 0,95 = 0,243$. For a more clinically interpretable evaluation, an inference pipeline with post-processing based on Extra NMS was implemented to remove highly overlapping duplicate detections, and two complementary metrics were computed: (i) the Jaccard Index (IoU) to quantify spatial agreement between predictions and ground truth, and (ii) an FROC curve (sensitivity vs. FPPI) to analyze the trade-off between sensitivity and false positives per image. On the test set, 39,61 % of the images exhibited $\text{IoU} \geq 0,60$ (143 of 361), and the FROC analysis reported sensitivities of 0.5435 at $\text{FPPI} \leq 0,5$, 0.6332 at $\text{FPPI} \leq 1,0$, and 0.6860 at $\text{FPPI} \leq 2,0$. These results suggest that the proposed approach can localize masses with significant spatial concordance in a relevant fraction of cases, providing a reproducible basis for clinical decision support systems.

Keywords: mammography, object detection, YOLOv8, CBIS-DDSM, Jaccard index, FROC, FPPI.

Índice general

1. Descripción del Problema	11
1.1. Planteamiento del Problema	11
1.1.1. Formulación	11
1.1.2. Sistematización	11
1.2. Objetivos	12
1.2.1. Objetivo General	12
1.2.2. Objetivos Específicos	12
1.3. Justificación	12
1.4. Delimitaciones y Alcances	13
1.4.1. Entregables	13
2. Marco Teórico	15
2.1. Introducción conceptual	15
2.1.1. Cáncer de mama y rol de la mamografía	15
2.1.2. Términos básicos en mamografía	15
2.1.3. Hallazgos imagenológicos y BI-RADS	16
2.2. Formación de imagen y características relevantes	16
2.2.1. Física básica de la mamografía digital	16
2.2.2. Apariencia de las masas y desafíos computacionales	17
2.3. Conjunto de datos CBIS-DDSM	17
2.3.1. Origen y propósito	17
2.3.2. Resolución espacial de CBIS-DDSM	17
2.3.3. Estructura de los casos de masas	18
2.4. De ROI mask a <i>bounding boxes</i> en formato YOLO	18
2.4.1. ¿Qué es YOLO?	18
2.4.2. Definiciones y notación	18
2.4.3. Conversión a formato YOLO (normalizado)	19
2.4.4. Conexión con el entrenamiento y la evaluación	19
2.5. Detección de objetos: panorama	19
2.5.1. Enfoques de dos etapas vs. una etapa	19
2.5.2. Trade-offs típicos	20
2.5.3. Elección práctica para mamografía: ¿por qué YOLO?	20
2.6. Arquitectura y objetivos de entrenamiento en YOLO	21
2.6.1. <i>Backbone</i> : extracción de características	21
2.6.2. <i>Neck</i> : agregación multi-escala (FPN/PAN)	21
2.6.3. <i>Head</i> : predicción densa y salidas	21
2.6.4. Pérdida de entrenamiento (<i>loss</i>): componentes	22

2.6.5.	Asignación de positivos/negativos	22
2.6.6.	Supresión de no-máximos (NMS) e inferencia	23
2.6.7.	Optimización y regularización prácticas	23
2.6.8.	Conexión con las métricas reportadas	23
2.7.	Métricas de evaluación	23
2.7.1.	Coincidencia por IoU y contabilidad de aciertos	23
2.7.2.	Precisión y <i>recall</i>	24
2.7.3.	AP y mAP	24
2.7.4.	F1-score y selección de umbrales	24
2.7.5.	Métricas complementarias para imagen médica	24
2.8.	Desafíos específicos del dominio	25
2.8.1.	Desbalance de clases y tamaño del objeto	25
2.8.2.	Resumen para este trabajo	25
2.9.	Trabajo relacionado	25
2.9.1.	Detectores en mamografía: two-stage vs. one-stage	25
2.9.2.	Segmentación vs. detección de masas	26
2.9.3.	Uso de CBIS-DDSM y colecciones cercanas	26
2.9.4.	Resumen de enfoques y conexión con este trabajo	26
3.	Metodología	29
3.1.	Preprocesamiento	29
3.1.1.	Visión general del pipeline	29
3.1.2.	Conjunto de datos	31
3.1.3.	Análisis exploratorio de metadatos de masas	31
3.1.4.	Implementación de funciones para cargue de imágenes	31
3.1.5.	Análisis exploratorio de imágenes	32
3.1.6.	Cálculo del centro de masa en las ROI Mask	36
3.1.7.	Generación y ajuste de bounding boxes	37
3.1.8.	Generación de la estructura de carpetas para YOLO	42
3.1.9.	Creación de etiquetas en formato YOLO	42
3.2.	Entrenamiento	44
3.2.1.	Entrenamiento del modelo YOLO	44
3.2.2.	Modelo base: YOLOv8m	44
3.2.3.	Entrenamiento del modelo y ajuste de hiperparámetros	45
3.2.4.	Evidencia visual del entrenamiento	46
4.	Resultados	51
4.1.	Selección final del modelo y evaluación clínica	51
4.1.1.	Motivación de la evaluación adicional	51
4.1.2.	Pipeline de inferencia y post-procesamiento	51
4.1.3.	Evidencia visual del post-procesamiento (Extra NMS)	52

Índice general **9**

4.1.4. Resultados con Índice de Jaccard	53
4.1.5. Resultados con FROC (Sensitivity vs FPPI)	55
5. Conclusiones	57
5.1. Limitaciones	58
5.2. Trabajo futuro	58
5.3. Cierre	58
Bibliografía	61

Descripción del Problema

1.1. Planteamiento del Problema

Durante las últimas décadas el cáncer de mama se ha convertido en uno de los tipos de cáncer con mayor incidencia y mortalidad anual. Esto se debe, en parte, a que puede presentarse en mujeres de distintas edades y, además, a la dificultad de detectar de manera oportuna o con suficiente certeza la aparición temprana de malformaciones o anomalías en la zona mamaria. Entre las alteraciones más comunes, pero potencialmente más riesgosas, se encuentran las masas o bultos en el seno. Estas pueden originarse por diversas causas y, en muchos casos, no representan un riesgo grave para la salud. No obstante, especialistas alrededor del mundo recomiendan abordar todas las masas con la debida importancia, dado que algunas lesiones inicialmente benignas pueden evolucionar hacia patologías malignas.

En este contexto, la mamografía constituye uno de los métodos más utilizados para la detección y el análisis de este tipo de hallazgos, apoyando al personal médico en la identificación temprana de masas u otras anomalías y, con ello, en la toma de decisiones clínicas basadas en la evidencia obtenida. Sin embargo, las mamografías digitales son imágenes de alta complejidad: incluso para un experto, la inspección visual y la localización precisa de masas puede resultar laboriosa y susceptible a variaciones. Como consecuencia, los costos asociados en tiempo y recursos para realizar la detección manual pueden ser elevados, lo cual puede reflejarse en retrasos en el proceso diagnóstico y en limitaciones de acceso a estos estudios.

1.1.1. Formulación

¿Cómo se podría detectar de manera automática la presencia de masas en mamografías digitales utilizando técnicas de inteligencia artificial?

1.1.2. Sistematización

- ¿Cuál será el conjunto de datos del que se obtendrán las mamografías digitales necesarias?
- ¿Qué atributos o características de los datos se tendrán en cuenta para entrenar el modelo?
- ¿Como se preprocesarán las mamografías digitales para facilitar la detección de masas?
- ¿Cuáles técnicas de aprendizaje automático se tendrán en cuenta para realizar la detección de masas en mamografías digitales?

- ¿De qué manera se evaluará y comparará el comportamiento del modelo desarrollado teniendo en cuenta los resultados obtenidos con base en el subconjunto de datos determinado para evaluación?

1.2. Objetivos

1.2.1. Objetivo General

Desarrollar un sistema de detección automática de masas en mamografías digitales.

1.2.2. Objetivos Específicos

- Seleccionar un conjunto de datos que contengan mamografías digitales.
- Identificar variables o atributos que ayudaran al entrenamiento del modelo.
- Implementar un algoritmo que se encargue de preprocesar las mamografías digitales para facilitar la detección de masas en estas.
- Seleccionar una técnica de deep learning de aprendizaje automático de máquina para implementar los modelos de detección de masas en mamografías digitales.
- Evaluar el comportamiento de cada modelo considerando el subconjunto de datos determinado para evaluación.

1.3. Justificación

Actualmente, la mamografía es uno de los métodos más utilizados para identificar de forma temprana posibles signos de cáncer de mama. Sin embargo, estos estudios no siempre son completamente asequibles o oportunos, debido a la complejidad de las imágenes y a la necesidad de personal especializado para su lectura e interpretación. Incluso para expertos, el análisis puede requerir un tiempo considerable, dado que primero deben reconocer la anatomía y luego identificar posibles masas o anomalías para proceder con su evaluación. En consecuencia, este proceso puede implicar un costo monetario elevado.

Teniendo esto en cuenta, un sistema capaz de apoyar y agilizar parte de este procedimiento podría contribuir a obtener resultados más tempranos y a reducir la carga de trabajo para los especialistas, ya sea en escenarios de tamizaje mediante mamografía o en procesos clínicos posteriores que requieran intervención.

Por lo anterior, este proyecto busca facilitar la tarea de detectar masas en mamografías digitales mediante técnicas de inteligencia artificial. De esta manera, se podría reducir el tiempo y esfuerzo invertido en la detección manual, generando beneficios tanto para las pacientes como para el personal médico. Además, este trabajo puede servir como base para investigaciones futuras orientadas a

automatizar procedimientos complejos en el ámbito de la salud, promoviendo un aprovechamiento responsable de técnicas modernas de aprendizaje automático.

Finalmente, se trata de un proyecto cuyo desarrollo depende principalmente de la disponibilidad de datos adecuados (mamografías digitales con anotaciones) y de recursos computacionales suficientes para el entrenamiento y evaluación de los modelos. Adicionalmente, su pertinencia en el ámbito médico se respalda en el hecho de que ya existen iniciativas similares en funcionamiento, como ScanNav, un sistema de detección en tiempo real por medio de imágenes de ultrasonido [1]. De este modo, con un desarrollo apropiado y una fase rigurosa de validación, un sistema de detección automática de masas en mamografías podría convertirse en una herramienta de apoyo útil en el entorno clínico.

1.4. Delimitaciones y Alcances

- Se emplearán métodos de aprendizaje profundo.
- Se trabajará con la base de datos CBIS-DDSM (Curated Breast Imaging Subset of Digital Database for Screening Mammography).
- El conjunto de datos a utilizar tiene un tamaño aproximado de 5 GB.
- Las imágenes del conjunto de datos se encuentran en formato DICOM.
- Se utilizarán máquinas virtuales proporcionadas por servicios de terceros o se hará uso de las máquinas físicas de la Universidad Pontificia Javeriana Cali.

1.4.1. Entregables

- Documento con la información relativa a las técnicas de aprendizaje automático utilizadas y sus respectivos resultados.
- Código fuente de los modelos implementados.

Marco Teórico

2.1. Introducción conceptual

La detección temprana del cáncer de mama es un problema de alto impacto sanitario, en el cual la mamografía juega un papel central como técnica de tamizaje y diagnóstico por imagen. En términos computacionales, las masas mamográficas constituyen un tipo de hallazgo relevante, caracterizado por regiones de mayor densidad con morfologías y márgenes que pueden variar ampliamente, lo que dificulta su identificación automática en imágenes con bajo contraste y superposición de tejidos. Esta sección introduce el contexto clínico mínimo, la terminología básica y el marco descriptivo de BI-RADS necesarios para situar el problema y motivar el uso de técnicas de aprendizaje automático en la detección de masas [2, 3].

2.1.1. Cáncer de mama y rol de la mamografía

El cáncer de mama se mantiene como una de las principales causas de incidencia y mortalidad por cáncer en mujeres a nivel global. La estrategia de tamizaje mediante mamografía permite detectar anomalías en etapas más tempranas, lo que se asocia con mejores desenlaces clínicos y opciones terapéuticas menos invasivas. Desde la perspectiva de procesamiento de señales e imágenes, esto plantea un problema de detección de patrones morfológicos sutiles inmersos en condiciones de ruido, bajo contraste y variabilidad anatómica. En consecuencia, los métodos computacionales buscan incrementar la sensibilidad (capacidad de detectar masas verdaderas) sin un aumento prohibitivo en los falsos positivos, manteniendo así la utilidad clínica del flujo de trabajo radiológico [2].

2.1.2. Términos básicos en mamografía

La mamografía digital produce imágenes en dos proyecciones estándar por mama: craneocaudal (CC) y mediolateral oblicua (MLO). Estas vistas capturan estructuras como el tejido fibroglandular, el músculo pectoral (especialmente visible en MLO) y el tejido adiposo, además de posibles artefactos de adquisición. Tales elementos son importantes para comprender la distribución de intensidades y la textura de fondo contra la cual deben identificarse las masas. Desde el punto de vista físico, la formación de la imagen depende de parámetros del sistema (espectro del tubo de rayos X, filtración, compresión, detector) que determinan la relación señal-ruido y el contraste, factores críticos para el desempeño de algoritmos de realce, segmentación y detección [4, 5]. En particular, la variabilidad entre dispositivos y protocolos conlleva un cambio de dominio que los modelos computacionales deben manejar adecuadamente.

2.1.3. Hallazgos imagenológicos y BI-RADS

BI-RADS (Breast Imaging Reporting and Data System) provee un léxico estandarizado para describir hallazgos y un sistema de categorías que comunica el nivel de sospecha de malignidad. En el caso de las masas, el léxico BI-RADS enfatiza atributos como forma (redonda, oval, irregular), márgenes (circunscritos, microlobulados, indistintos, espiculados, oscurecidos) y densidad relativa (alta, igual, baja), que guían la valoración clínica. Aunque en esta tesis el objetivo no es la evaluación diagnóstica completa ni la asignación de categorías BI-RADS, la estandarización terminológica es valiosa para: (i) precisar qué se entiende por masa en el contexto de la imagen, (ii) motivar las dificultades asociadas a márgenes poco definidos o espiculados, y (iii) alinear la detección automática con el objetivo práctico de resaltar regiones susceptibles de análisis por parte del radiólogo [3]. En términos computacionales, estos descriptores se relacionan con patrones de borde, textura y contraste; de allí que la detección de masas requiera algoritmos robustos a variaciones de forma y a bordes con gradientes suaves, preservando la especificidad para minimizar el número de hallazgos falsos que interrumpen el flujo de lectura.

2.2. Formación de imagen y características relevantes

La mamografía digital registra diferencias sutiles de atenuación de rayos X entre tejidos mamaros. Desde una perspectiva de ingeniería, la calidad de la imagen depende de la interacción radiación–materia (atenuación), del sistema de adquisición (espectro del tubo, filtración, compresión), y de la respuesta del detector digital. Estos factores determinan la relación señal-ruido (SNR) y el contraste, que condicionan tanto la percepción humana como el desempeño de algoritmos de detección automatizada [5, 4].

2.2.1. Física básica de la mamografía digital

En mamografía, el haz de rayos X se atenúa al atravesar el tejido según la ley de Beer–Lambert. En el dominio digital, los detectores de panel plano (p.ej., a-Si con centelleador CsI:Tl) convierten esa señal en una imagen discreta $I \in \mathbb{R}^{H \times W}$ con cuantificación en bits. El fabricante implementa procesos de corrección de campo plano (*ganancia* y *dark current*) para reducir no-uniformidades del detector; no obstante, persisten fuentes de ruido cuántico y electrónico que limitan la SNR efectiva [5]. La compresión mamaria aumenta la uniformidad del espesor y reduce movimiento, mejorando SNR y contraste.

Para caracterizar el contraste local entre una región sospechosa (masa) y el fondo, es útil una medida adimensional:

$$C = \frac{\mu_{\text{masa}} - \mu_{\text{fondo}}}{\sigma_{\text{fondo}}}, \quad (2.1)$$

donde μ y σ son la media y desviación estándar de niveles de gris en ventanas locales. Masas de bajo contraste ($C \approx 1$) son especialmente desafiantes, pues sus bordes presentan gradientes suaves y se confunden con textura fibroglandular [4].

2.2.2. Apariencia de las masas y desafíos computacionales

En mamografía, las masas suelen ser hiperdensas respecto al tejido adiposo y pueden exhibir morfologías redondas, ovales o irregulares, con márgenes que van de circunscritos a espiculados (puntas radiadas). Los bordes espiculados implican transiciones de intensidad graduales y estructuras lineales finas alrededor del nódulo, lo cual dificulta la segmentación basada en umbrales y vuelve crucial la explotación de textura y contexto multiescala. Además, la superposición de tejido (especialmente en proyección MLO) puede ocultar o mimetizar lesiones, incrementando falsos negativos y positivos.

Para métodos de detección tipo YOLO, estas propiedades se traducen en tres retos: (i) variabilidad intra/inter-lesión de tamaño y forma (necesidad de *receptive fields* multiescala), (ii) bajo contraste local (beneficio potencial de realce previo), y (iii) textura de fondo compleja (necesidad de *features* robustas y aumento de datos realista) [6].

2.3. Conjunto de datos CBIS-DDSM

El *Curated Breast Imaging Subset of DDSM* (CBIS-DDSM) es una versión curada y estandarizada del histórico *Digital Database for Screening Mammography* (DDSM), diseñada para facilitar la investigación en análisis de imágenes mamográficas. CBIS-DDSM ofrece casos anotados con segmentaciones de lesiones y metadatos radiológicos, lo que permite tareas de clasificación, segmentación y detección de hallazgos como masas y calcificaciones [7, 8, 9].

2.3.1. Origen y propósito

DDSM fue una de las primeras colecciones públicas de mamografías digitalizadas, con anotaciones clínicas y de expertos [8]. CBIS-DDSM retoma ese acervo, corrige inconsistencias, homogeneiza estructuras de directorio y entrega archivos de imagen y máscaras en formatos accesibles para flujos de trabajo modernos, además de tablas con metadatos por estudio/lesión [7, 9]. El objetivo principal de CBIS-DDSM es proveer un conjunto reproducible que reduzca fricción de preprocesamiento y permita comparar algoritmos de aprendizaje automático bajo supuestos similares.

2.3.2. Resolución espacial de CBIS-DDSM

El conjunto CBIS-DDSM proviene de mamografías en película digitalizadas (*scanned film*), por lo que la resolución espacial depende del digitalizador utilizado. En la literatura se reportan tamaños de píxel típicos de $42 \mu\text{m}/\text{píxel}$, $43,5 \mu\text{m}/\text{píxel}$ y $50 \mu\text{m}/\text{píxel}$ (equivalentes a $0,042 \text{ mm}/\text{píxel}$, $0,0435 \text{ mm}/\text{píxel}$ y $0,050 \text{ mm}/\text{píxel}$), asociados a distintos digitalizadores empleados en DDSM/CBIS-DDSM. Estos valores pueden variar entre estudios, por lo que el espaciado puede verificarse directamente desde el metadato DICOM (`PixelSpacing`), el cual define la distancia física entre centros de píxel (en mm). [10, 11, 12, 13]

2.3.3. Estructura de los casos de masas

Para el subproblema de detección de masas (ignorando calcificaciones), cada caso suele incluir:

- **Imágenes mamográficas** por mama y proyección estándar: craneocaudal (CC) y mediolateral oblicua (MLO). Las resoluciones varían por caso y equipo de adquisición.
- **Máscaras de región de interés (ROI)** por lesión, provistas como imágenes binarias (o derivables a partir de contornos), que segmentan el área de la masa.
- **Metadatos** a nivel de estudio/imagen/lesión: identificador de paciente, lateralidad (L/R), proyección (CC/MLO), información patológica (benigno/maligno cuando está disponible), y referencias a los archivos de imagen y ROI.

En la práctica, una imagen puede contener múltiples masas; en tal caso, CBIS-DDSM lista varias ROI asociadas a esa misma vista.

2.4. De ROI mask a *bounding boxes* en formato YOLO

2.4.1. ¿Qué es YOLO?

YOLO (*You Only Look Once*) es una familia de detectores de objetos **one-stage** que formula la detección como una tarea de regresión directa sobre cuadrículas (o puntos de referencia) y múltiples escalas, prediciendo para cada ubicación: cajas delimitadoras, puntajes de confianza y probabilidades de clase en una sola pasada por la red [14]. Frente a detectores **two-stage** (p. ej., R-CNN), YOLO privilegia eficiencia y latencia baja, con un compromiso de precisión competitivo. Las variantes modernas (p. ej., YOLOv4, YOLOv8) incorporan *backbones* y *necks* con agregación multiescala (FPN/PAN), pérdidas basadas en IoU (GIoU/CIoU) y técnicas de entrenamiento robustas (augmentaciones, *label assignment* dinámico) [15, 6, 16, 17].

En el contexto de mamografía:

- El carácter multiescala de YOLO es útil para masas pequeñas y grandes en el mismo campo.
- La pérdida basada en IoU favorece una localización más estable bajo fondos complejos.
- La predicción densa facilita mantener sensibilidad alta (recall), clave en tamizaje.

En este trabajo de grado, se emplea YOLO con una **única clase (mass)** y etiquetas derivadas directamente de las máscaras de CBIS-DDSM.

2.4.2. Definiciones y notación

Sea la imagen original $I \in \mathbb{R}^{H_0 \times W_0}$ y la máscara binaria de la lesión $M \in \{0, 1\}^{H_0 \times W_0}$. Denotemos por

$$S = \{(x, y) \mid M(x, y) = 1\}$$

el soporte de la lesión. El *bounding box* mínimo que circunscribe S se define por:

$$x_{\min} = \min_{(x,y) \in S} x, \quad x_{\max} = \max_{(x,y) \in S} x, \quad y_{\min} = \min_{(x,y) \in S} y, \quad y_{\max} = \max_{(x,y) \in S} y.$$

Para control geométrico, es útil el centroide:

$$x_c = \frac{1}{|S|} \sum_{(x,y) \in S} x, \quad y_c = \frac{1}{|S|} \sum_{(x,y) \in S} y. \quad (2.2)$$

2.4.3. Conversión a formato YOLO (normalizado)

El formato YOLO exige centro y tamaño normalizados respecto de W, H :

$$x = \frac{\tilde{x}_{\min} + \tilde{x}_{\max}}{2W}, \quad y = \frac{\tilde{y}_{\min} + \tilde{y}_{\max}}{2H}, \quad w = \frac{\tilde{x}_{\max} - \tilde{x}_{\min}}{W}, \quad h = \frac{\tilde{y}_{\max} - \tilde{y}_{\min}}{H}, \quad (2.3)$$

con $x, y, w, h \in (0, 1]$. La etiqueta final se escribe como:

`class x y w h`

donde `class` = 0 para `mass`. Si una imagen contiene K masas, se generan K líneas (una por caja). Es fundamental aplicar la misma transformación geométrica (escala y *padding*) a imagen y máscara para que las coordenadas sean coherentes.

2.4.4. Conexión con el entrenamiento y la evaluación

Etiquetas coherentes reducen errores sistemáticos en *box loss* y estabilizan la asignación de positivos/negativos. Durante inferencia, las predicciones se filtran con NMS (*Non-Maximum Suppression*) para suprimir duplicados; variantes como *Soft-NMS* pueden mejorar la retención de verdaderos positivos cuando existen detecciones cercanas [18]. Las métricas de reporte siguen el estándar en YOLO: Precision, Recall, AP@0,5 y mAP@[0,5 : 0,95]. En mamografía, priorizar el *recall* (sensibilidad) es clave; por ello, es recomendable ajustar umbrales de confianza y NMS en validación para equilibrar sensibilidad y tasa de falsos positivos, con análisis estratificado por tamaño de masa y por vista (CC/MLO).

2.5. Detección de objetos: panorama

2.5.1. Enfoques de dos etapas vs. una etapa

Los detectores de objetos modernos se organizan clásicamente en dos familias:

- **Dos etapas (two-stage):** primero generan propuestas de regiones (posibles objetos) y luego refinan clasificación y regresión de cajas sobre esas propuestas. Los representantes clásicos son R-CNN, Fast R-CNN y Faster R-CNN, este último integrando una *Region Proposal Network* (RPN) que acelera la generación de propuestas [19, 20, 21]. Suelen ofrecer alta precisión, especialmente en objetos pequeños, a costa de mayor complejidad y latencia.

- **Una etapa (one-stage):** formulan la detección como regresión y clasificación densas sobre una rejilla de ubicaciones o puntos de referencia, sin etapa explícita de propuestas. Ejemplos: SSD, RetinaNet y la familia YOLO [22, 23, 14]. Se destacan por su eficiencia (tiempo de inferencia bajo) y alto rendimiento con entrenamiento cuidadoso.

2.5.2. Trade-offs típicos

Velocidad vs. precisión. Los métodos one-stage suelen ser más rápidos que los two-stage, lo que los hace atractivos para escenarios con restricciones de tiempo o recursos. En evaluación estandarizada (p. ej., COCO) los two-stage históricamente lograron ventajas en precisión para objetos pequeños, pero las mejoras en *neck* multiescala (FPN/PAN), asignación de positivos y pérdidas basadas en IoU han cerrado esa brecha en varios dominios [23, 15, 6].

Clases, anchors y multiescala. Muchos detectores usan *anchors* (cajas base) de distintas escalas y relaciones de aspecto por nivel de la pirámide. SSD/RetinaNet y YOLO (en varias versiones) explotan esta idea; variantes más recientes también exploran enfoques *anchor-free*. En imágenes médicas, donde el número de clases es pequeño (una clase en este trabajo) y la variabilidad de tamaño es alta, las pirámides de características ayudan a retener sensibilidad para masas pequeñas sin perder desempeño en lesiones grandes.

Desbalance positivo/negativo. En detección densa hay abundantes negativos (fondo) y pocos positivos. RetinaNet introdujo la *Focal Loss* para atenuar la contribución de ejemplos fáciles y concentrar el aprendizaje en los difíciles:

$$\mathcal{L}_{\text{focal}} = -\alpha (1 - p_t)^\gamma \log(p_t), \quad p_t = \begin{cases} p & \text{si la clase es positiva} \\ 1 - p & \text{si la clase es negativa} \end{cases} \quad (2.4)$$

con $\gamma > 0$ y $\alpha \in (0, 1)$ [23]. En YOLO, el desbalance se maneja combinando *objectness*, BCE focalizada y esquemas de asignación (p. ej., por IoU/DFL).

2.5.3. Elección práctica para mamografía: ¿por qué YOLO?

En el problema de detectar masas en mamografías digitalizadas (una sola clase, tamaños variables, bajo contraste):

1. **Eficiencia y estabilidad:** YOLO ofrece inferencia rápida y estable, útil para iterar en *validation* con distintos umbrales y ajustes de NMS sin tiempos de cómputo elevados.
2. **Multiescala efectiva:** Las cabeceras de detección en múltiples resoluciones (*FPN/PAN*) ayudan a capturar masas pequeñas (difíciles) y masas grandes en el mismo marco.
3. **Pérdidas basadas en IoU y DFL:** El uso de GIoU/CIoU y *Distribution Focal Loss* (en variantes modernas) mejora la calidad de la regresión de cajas, importante cuando los márgenes son difusos.

4. **Simplicidad de integración:** Dado que las etiquetas se derivan directamente de las ROI masks (una línea `class x y w h` por masa), el flujo ROI→bbox YOLO es natural y reproducible, reduciendo errores de ingeniería.

En este trabajo, el *pipeline* queda así encadenado: **CBIS-DDSM** → **ROI binaria** → **bbbox normalizada (YOLO)** → **entrenamiento de un detector one-stage (YOLO)**. Esta decisión busca maximizar *recall* con una tasa de falsos positivos controlada, alineando el objetivo computacional con la prioridad clínica de no perder masas verdaderas.

2.6. Arquitectura y objetivos de entrenamiento en YOLO

En la familia YOLO, la red se organiza en tres componentes: *backbone* (extracción de características), *neck* (agregación multi-escala) y *head* (predicción de cajas, confianzas y clases). Las variantes modernas combinan pirámides de características (FPN/PAN), pérdidas basadas en IoU para la regresión de cajas y mecanismos de asignación de positivos/negativos robustos [24, 25, 6].

2.6.1. Backbone: extracción de características

El *backbone* transforma la imagen normalizada $I \in \mathbb{R}^{H \times W}$ en mapas de características a múltiples resoluciones mediante convoluciones, activaciones no lineales (p. ej., SiLU) y bloques con atajos residuales. Las primeras capas capturan bordes y texturas finas (útiles para márgenes de masas); las profundas agregan contexto semántico (patrones de tejido). La salida es un conjunto de *feature maps* piramidales $\{P_3, P_4, P_5, \dots\}$ con tamaños decrecientes y mayor campo receptivo.

2.6.2. Neck: agregación multi-escala (FPN/PAN)

Para retener sensibilidad a objetos pequeños y grandes en la misma escena, YOLO emplea una pirámide de características con rutas de arriba-abajo (FPN) y abajo-arriba (PANet). FPN fusiona semántica fuerte desde niveles profundos hacia resoluciones altas; PAN refuerza información espacial ascendente, mejorando la localización precisa [24, 25]. El resultado son mapas enriquecidos a varias escalas sobre los que la *head* emite predicciones densas.

2.6.3. Head: predicción densa y salidas

La *head* produce, por celda y por escala, un vector con:

- Parámetros de caja $(\hat{x}, \hat{y}, \hat{w}, \hat{h})$ (en variantes *anchor-based* o *anchor-free*).
- *Objectness* $\hat{o} \in [0, 1]$ (probabilidad de que exista un objeto).
- Probabilidades de clase $\hat{\mathbf{p}}$ (en este trabajo, una clase: `mass`).

En inferencia, se combinan \hat{o} y $\hat{\mathbf{p}}$ para puntuar cada predicción y se aplica supresión de no-máximos (NMS) para eliminar solapamientos redundantes.

2.6.4. Pérdida de entrenamiento (*loss*): componentes

La función objetivo típica suma tres términos:

$$\mathcal{L} = \lambda_{\text{box}} \mathcal{L}_{\text{box}} + \lambda_{\text{obj}} \mathcal{L}_{\text{obj}} + \lambda_{\text{cls}} \mathcal{L}_{\text{cls}}.$$

Regresión de cajas (\mathcal{L}_{box}). Se utilizan pérdidas basadas en IoU. Sea B la caja predicha y B^* la de verdad de terreno. La intersección sobre unión es:

$$\text{IoU}(B, B^*) = \frac{|B \cap B^*|}{|B \cup B^*|}.$$

GIoU mejora la señal cuando no hay intersección, penalizando el área ociosa del rectángulo mínimo que contiene a ambas cajas C :

$$\text{GIoU} = \text{IoU} - \frac{|C \setminus (B \cup B^*)|}{|C|}, \quad \mathcal{L}_{\text{GIoU}} = 1 - \text{GIoU} [16].$$

CIoU añade términos de distancia entre centros y consistencia de aspecto:

$$\mathcal{L}_{\text{CIoU}} = 1 - \text{IoU} + \frac{\rho^2(\mathbf{c}, \mathbf{c}^*)}{d^2} + \alpha v,$$

donde ρ es la distancia euclídea entre centros, d la diagonal de C , v mide la diferencia de razón de aspecto y α es un factor de equilibrio [17]. Variantes recientes incorporan *Distribution Focal Loss* (DFL) para refinar los bordes tratando cada coordenada como una distribución discreta sobre bins, logrando regresión sub-píxel [26].

Objectness y clasificación ($\mathcal{L}_{\text{obj}}, \mathcal{L}_{\text{cls}}$). Se usan pérdidas logísticas (BCE) y, en escenarios con fuerte desbalance, variantes focales que atenúan ejemplos fáciles:

$$\mathcal{L}_{\text{focal}} = -\alpha(1 - p_t)^\gamma \log(p_t),$$

con $\gamma > 0$ y $\alpha \in (0, 1)$ [23]. En este trabajo, al haber una sola clase, \mathcal{L}_{cls} se reduce a una BCE binaria por el canal *mass*. El *objectness* actúa como filtro de fondo, clave para controlar falsos positivos en tejido complejo.

2.6.5. Asignación de positivos/negativos

El entrenamiento requiere decidir qué predicciones (por celda/escala) son positivas para una caja GT. Hay dos familias:

- **Asignación por umbrales de IoU:** una predicción es positiva si su IoU con la GT supera un umbral (p. ej., 0.5) y negativa si está por debajo de otro (p. ej., 0.4), dejando una zona gris.
- **Asignación dinámica:** esquemas como OTA/SimOTA ponderan distancia/IoU y puntajes para seleccionar un conjunto de candidatos por GT, estabilizando el aprendizaje y mejorando el *recall* [27].

En cualquiera de los dos casos, los negativos dominan numéricamente; por eso se usan *objectness* y/o pérdidas focales.

2.6.6. Supresión de no-máximos (NMS) e inferencia

Al puntuar cada predicción con $s = \hat{o} \cdot \hat{p}_{\text{mass}}$ y filtrar por un umbral de confianza, NMS elimina cajas con alta superposición conservando la de mayor puntaje. *Soft-NMS* reduce el puntaje de las cajas solapadas en vez de descartarlas, lo que puede ayudar cuando hay detecciones cercanas de la misma masa [18]. En validación, los umbrales de confianza y NMS se ajustan para equilibrar *recall* y *precision*.

2.6.7. Optimización y regularización prácticas

En la práctica se usan optimizadores como SGD con momentum o AdamW, *weight decay*, y programadores de LR (cosine/step). Las estrategias de aumento (p. ej., flips, rotaciones moderadas, jitter de brillo/contraste) deben preservar la semántica clínica (márgenes y forma de las masas). El *early stopping* con paciencia basada en mAP o *val loss* evita sobreajuste, especialmente con splits por paciente.

2.6.8. Conexión con las métricas reportadas

La calidad de la regresión (GIoU/CIoU/DFL) y una asignación de positivos estable se reflejan en:

- **mAP@0.5**: sensibilidad a errores moderados de localización (útil para comparar con literatura de mamografía).
- **mAP@[0.5:0.95]**: penaliza más los errores de caja, alineado con CIoU/DFL.
- **Precision/Recall**: se ajustan con el umbral de confianza y NMS; en detección de masas, priorizar *recall* es clínicamente razonable, con monitorización de la tasa de falsos positivos por imagen.

En este trabajo (una sola clase), simplificar la cabeza de clasificación y sintonizar λ_{obj} ayuda a mantener *recall* alto sin degradar en exceso la *precision*.

2.7. Métricas de evaluación

La evaluación en detección de objetos cuantifica la localización y la decisión (presencia de objeto) bajo umbrales de confianza y de superposición geométrica. En este trabajo se reportan Precision, Recall, AP@0,5, mAP@[0,5:0,95] y, como análisis complementario clínico, la tasa de falsos positivos por imagen y curvas FROC.

2.7.1. Coincidencia por IoU y contabilidad de aciertos

Dada una predicción B y una verdad de terreno B^* , la intersección sobre unión (IoU) es:

$$\text{IoU}(B, B^*) = \frac{|B \cap B^*|}{|B \cup B^*|} \in [0, 1]. \quad (2.5)$$

Para un umbral τ (p. ej., $\tau = 0,5$), una predicción es verdadero positivo (TP) si su IoU con alguna B^* no asignada es $\geq \tau$. Si no coincide con ninguna B^* , es falso positivo (FP). Cualquier B^* no emparejada tras procesar todas las predicciones es un falso negativo (FN). El emparejamiento se realiza ordenando las predicciones por puntaje (confianza) y aplicando supresión de no-máximos (NMS) para evitar duplicados.

2.7.2. Precisión y *recall*

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}. \quad (2.6)$$

Al variar el umbral de confianza se obtiene la curva Precisión-*Recall* (PR). En detección de masas, es habitual priorizar *recall* (sensibilidad), con control de FP mediante NMS y el umbral de confianza.

2.7.3. AP y mAP

El **Average Precision** (AP) es el área bajo la curva PR para una clase dada y un umbral de IoU fijo. Para AP@0,5, se fija $\tau = 0,5$. En **COCO**, la mAP@[0,5;0,95] promedia el AP en once umbrales $\tau \in \{0,50, 0,55, \dots, 0,95\}$ (muestreados típicamente en 101 puntos de *recall*) [28]. Formalmente, para una clase c :

$$\text{AP}_{c,\tau} = \int_0^1 \text{Precision}_{c,\tau}(r) dr, \quad \text{mAP} = \frac{1}{|\mathcal{T}|} \sum_{\tau \in \mathcal{T}} \left(\frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \text{AP}_{c,\tau} \right), \quad (2.7)$$

donde \mathcal{C} es el conjunto de clases y \mathcal{T} el conjunto de umbrales de IoU. En este trabajo, $|\mathcal{C}| = 1$ (clase *mass*), por lo que mAP coincide con el AP promedio en los umbrales \mathcal{T} .

2.7.4. F1-score y selección de umbrales

Para seleccionar umbrales de despliegue (confianza y NMS), puede optimizarse la métrica F1 en validación:

$$F_1 = 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (2.8)$$

En mamografía, priorizar *Recall* suele ser clínicamente preferible; por ello, también es común reportar F_β con $\beta > 1$ (p. ej., $\beta = 2$) para ponderar más la sensibilidad.

2.7.5. Métricas complementarias para imagen médica

Además de AP/mAP, en lectura clínica interesa la tasa de falsos positivos por imagen (FP/imagen), especialmente en escenarios de tamizaje, y las curvas **FROC** (Free-response ROC), que grafican *sensitivity vs. FP por imagen* [29, 30]. En FROC, un punto de operación con *sensitivity* = 0,9 y 0,5 FP/imagen puede ser preferible a otro con similar sensibilidad pero 2 FP/imagen, por su menor carga de *callbacks*. Para evaluar equidad del modelo, es útil estratificar resultados por tamaño físico de la masa (si hay *pixel spacing*) y por vista (CC/MLO).

2.8. Desafíos específicos del dominio

La detección de masas en mamografías digitalizadas enfrenta retos que no siempre aparecen en conjuntos de imágenes naturales. Estos desafíos afectan tanto la calidad de los datos como la validez externa del modelo entrenado, así como su utilidad clínica. A continuación se describen los principales aspectos y las medidas de mitigación recomendadas para este trabajo.

2.8.1. Desbalance de clases y tamaño del objeto

En CBIS-DDSM el número de píxeles de fondo supera ampliamente al de lesión; incluso contando por instancias, las masas pequeñas son menos frecuentes que las medianas/grandes. Este desbalance induce gradientes dominados por el fondo, y dificulta aprender buenos predictores para masas pequeñas. En detectores densos, además, existe desbalance positivo/negativo por cada celda/escala. Para mitigarlo:

- Emplear pérdidas que atenúen ejemplos fáciles (p. ej., focal en \mathcal{L}_{obj} o \mathcal{L}_{cls}) y pérdidas de caja robustas (GIoU/CIoU/DFL).
- Aumentos que preserven semántica y favorezcan objetos pequeños (crops con *zoom-in* moderado, rotaciones suaves).
- Reporte estratificado por tamaño físico (si hay *pixel spacing*) y por vista (CC vs. MLO) para visibilizar fallos en el estrato difícil.

2.8.2. Resumen para este trabajo

En este proyecto (CBIS-DDSM \rightarrow ROI \rightarrow bbox YOLO \rightarrow entrenamiento), se mitigan los riesgos anteriores mediante: (i) normalización y *letterboxing* coherentes, (ii) split por paciente, (iii) pérdidas de caja robustas y aumentos prudentes orientados a masas pequeñas, (iv) ajuste de umbrales con reporte de FP/imagen y FROC, y (v) registro completo de configuraciones y resultados para asegurar reproducibilidad.

2.9. Trabajo relacionado

Esta sección sitúa el enfoque adoptado (ROI \rightarrow bbox YOLO) dentro del panorama de la literatura en mamografía. Se organiza por (i) detectores de una y dos etapas aplicados a masas, (ii) segmentación vs. detección, y (iii) uso de CBIS-DDSM y colecciones afines. El objetivo es justificar la elección de un detector one-stage (YOLO) con una única clase (**mass**) y enlazar con las métricas reportadas (Precision, Recall, AP, mAP).

2.9.1. Detectores en mamografía: two-stage vs. one-stage

La literatura temprana en detección de lesiones con *deep learning* en mamografía adoptó con frecuencia esquemas **two-stage** (p. ej., Faster R-CNN), con buenos resultados en sensibilidad a costa

de mayor latencia. Por ejemplo, ribli2018 emplean un detector basado en regiones para localizar lesiones y luego clasificarlas, mostrando mejoras frente a métodos clásicos de CAD. Paralelamente, la familia **one-stage** (SSD/YOLO) ganó tracción por su eficiencia y rendimiento competitivo en objetos pequeños cuando se combinan pirámides de características y pérdidas basadas en IoU [22, 14, 23]. En mamografía, donde hay una única clase de interés y gran variabilidad de tamaño, los detectores one-stage ofrecen una vía directa para optimizar sensibilidad y FP/imagen mediante ajuste de umbrales y NMS, lo que facilita la integración práctica en flujos de validación.

2.9.2. Segmentación vs. detección de masas

Otra línea relevante modela las masas como regiones segmentables, usualmente con arquitecturas tipo U-Net, a fin de obtener contornos detallados y medidas morfológicas [31]. Este enfoque favorece estudios que requieren *shape analysis* o márgenes finos (p. ej., espiculaciones), pero demanda máscaras de alta calidad para el entrenamiento y suele ser más costoso computacionalmente en validación extensiva. En contraste, la detección formula el problema como localización caja+confianza, suficiente para priorizar casos y calcular métricas estándar (AP/mAP, FROC). Híbridos como *Mask R-CNN* combinan ambas ideas con una *branch* de máscaras [32], pero incrementan complejidad. En tesis de pregrado, y con máscaras ROI disponibles (CBIS-DDSM), derivar *bounding boxes* y entrenar un detector one-stage ofrece una relación esfuerzo-beneficio favorable para estudiar sensibilidad y FP/imagen.

2.9.3. Uso de CBIS-DDSM y colecciones cercanas

CBIS-DDSM [7] ha sido ampliamente adoptado como banco de pruebas curado del histórico DDSM [8], facilitando comparaciones reproducibles y la disponibilidad de ROI para segmentación y/o detección. Trabajos de clasificación a nivel de estudio/imagen en screening (no necesariamente orientados a cajas) han mostrado que sistemas profundos bien entrenados pueden alcanzar desempeño competitivo con lectura clínica [33, 34], lo que subraya el potencial de *deep learning* en mamografía. Para detección de masas, CBIS-DDSM permite: (i) generar cajas a partir de máscaras (tu pipeline), (ii) evaluar mAP y FROC, y (iii) estratificar por vista (CC/MLO) y tamaño físico si hay *pixel spacing*. Este trabajo sigue esa línea, priorizando *recall* y controlando FP/imagen mediante NMS y umbrales.

2.9.4. Resumen de enfoques y conexión con este trabajo

- **Two-stage (Faster R-CNN):** alta precisión y buenos resultados en objetos pequeños; mayor complejidad y latencia [35, 21].
- **One-stage (SSD/YOLO):** eficiencia y rendimiento competitivo con FPN/PAN y pérdidas IoU/focal [22, 14, 23]. En una sola clase como *mass*, simplifica entrenamiento y tuning.
- **Segmentación (U-Net):** contornos precisos; requiere máscaras y mayor costo; útil como línea futura para análisis morfológico [31].

Dado que el objetivo de esta tesis es estudiar detección de masas optimizando sensibilidad y mAP con etiquetas `class x y w h` derivadas de ROI, el enfoque one-stage (YOLO) con preprocesamiento coherente (normalización+*letterboxing*) proporciona un marco sólido, reproducible y alineado con las métricas reportadas.

Metodología

3.1. Preprocesamiento

En este capítulo se describe detalladamente el procedimiento seguido para desarrollar el sistema de detección de masas en mamografías digitalizadas. La metodología se organiza en: (i) descripción de los datos utilizados, (ii) preprocesamiento y generación de etiquetas a partir de las ROI masks, (iii) configuración y entrenamiento del modelo YOLO, y (iv) diseño de los experimentos y métricas de evaluación.

3.1.1. Visión general del pipeline

En la Figura 3.1 se ilustra el flujo general de procesamiento implementado en este trabajo: desde la obtención de las imágenes de CBIS-DDSM hasta la evaluación del modelo de detección.

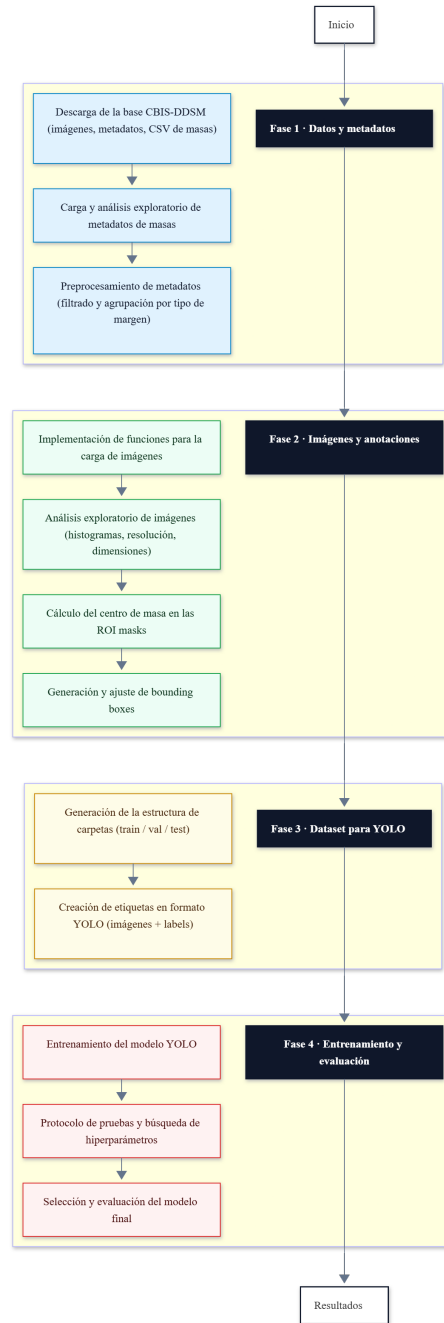


Figura 3.1: Esquema general del pipeline de detección de masas.

3.1.2. Conjunto de datos

CBIS-DDSM se organiza en dos subconjuntos principales: calcificaciones y masas. En esta investigación se utiliza exclusivamente el subconjunto correspondiente a masas (*Mass*), dado que el objetivo es la detección automática de este tipo de lesión en mamografías. El subconjunto empleado cuenta con aproximadamente 1710 masas anotadas en 1592 imágenes, provenientes de 892 pacientes.

Para la adquisición de los datos se utilizó la herramienta *NBIA Data Retriever*, mediante la cual se descargaron las mamografías y los archivos que contienen la información asociada a cada caso (metadatos y atributos por lesión). A partir de estos insumos se construyó el conjunto final de trabajo y se prepararon las anotaciones para el entrenamiento y evaluación del modelo, tal como se describe en las secciones posteriores.

3.1.3. Análisis exploratorio de metadatos de masas

Cada caso de masa contiene 14 atributos, variados entre atributos de tipo texto o de tipo numérico. En el Cuadro 3.1 se puede observar los diferentes atributos.

Aclaración sobre los conteos. Si bien el conjunto CBIS-DDSM (subconjunto de masas) cuenta con un total de **1710** casos de masa disponibles, en este trabajo el entrenamiento del modelo se realizó exclusivamente con los casos asignados a la partición de entrenamiento y validación. En particular, del total de casos, **1349** fueron destinados al desarrollo del modelo, de los cuales **1102** se utilizaron para entrenamiento y **247** para validación. Los casos restantes se reservaron para la evaluación final en el conjunto de prueba.

Teniendo la lista final de masas con las que se va a trabajar se procede a realizar ajustes finales para conectar bien cada masa con su respectiva imagen. Y así se pueden empezar a analizar las imágenes, en este caso, del conjunto de entrenamiento.

3.1.4. Implementación de funciones para cargue de imágenes

Las imágenes proporcionadas por el conjunto de datos, como es mencionado anteriormente, estaban codificadas en un formato médico llamado DICOM. Esto dificultó un poco la tarea de cargar y posteriormente analizar dichas imágenes, ya que no se podía por medio de metodologías tradicionales. Se requería de librerías como *pydicom* para poder cargar estas imágenes y acceder a sus metadatos.

Uno de los principales retos durante el cargue de imágenes fue la capacidad de memoria RAM para almacenarlas mientras eran analizadas en tiempo de ejecución, ya que se trataba de más de 1000 imágenes de alta resolución y en un formato ligeramente más pesado de lo normal. La estrategia utilizada consistió en simplemente cargar cada imagen en el momento justo en el que fuera a ser utilizada (para graficar, extraer datos, etc.), por lo tanto lo único que realmente se almacenaba era el file path. También se agruparon los filepaths según su número de ID del paciente, mama izquierda o derecha y tipo de vista.

Cuadro 3.1: Descripción de los atributos del subconjunto de masas del conjunto de datos CBIS-DDSM.

Atributo	Tipo	Descripción
patient_id	Cadena	Identificador anónimo del paciente al que pertenece el estudio mamográfico.
breast_density	Entero	Categoría de densidad mamaria según la escala BI-RADS (por ejemplo, de 1 a 4), que describe la proporción de tejido fibroglandular y tejido graso.
left or right breast	Cadena	Indica si la imagen corresponde a la mama izquierda (LEFT) o derecha (RIGHT).
image view	Cadena	Vista de la mamografía, típicamente CC (cráneo-caudal) o MLO (medio-lateral oblicua).
abnormality id	Entero	Identificador numérico de la anomalía dentro del estudio; permite distinguir múltiples lesiones en un mismo paciente.
abnormality type	Cadena	Tipo de anomalía anotada (por ejemplo, <i>mass</i> , <i>calcification</i>); en este trabajo se consideran únicamente las masas.
mass shape	Cadena	Forma de la masa (por ejemplo, <i>round</i> , <i>oval</i> , <i>lobular</i> , <i>irregular</i>), según la anotación del dataset.
mass margins	Cadena	Tipo de margen de la masa (por ejemplo, <i>circumscribed</i> , <i>obscured</i> , <i>microlobulated</i> , <i>ill-defined</i> , <i>spiculated</i>).
assessment	Entero	Evaluación numérica de la probabilidad de malignidad basada en la escala BI-RADS (valores más altos indican mayor sospecha).
pathology	Cadena	Resultado patológico asociado a la lesión (por ejemplo, BENIGN, MALIGNANT), según el diagnóstico de referencia.
subtlety	Entero	Grado de sutileza o dificultad de detección visual de la lesión en una escala ordinal.
image file path	Cadena	Ruta al archivo de la mamografía original dentro de la estructura de directorios del conjunto de datos.
cropped image file path	Cadena	Ruta al archivo de la imagen recortada que contiene la región alrededor de la anomalía.
ROI mask file path	Cadena	Ruta al archivo de la máscara binaria de la región de interés (ROI) que delimita la masa en la imagen recortada.

3.1.5. Análisis exploratorio de imágenes

Cada caso de estudio de mamografía contiene como mínimo 3 imágenes en formato DICOM, siendo la primera la mamografía digital completa (ya sea RIGHT o LEFT, CC o MLO), y después se proporciona una ROI Mask y una imagen recortada por cada masa presente en la mamografía. A continuación se puede evidenciar un ejemplo de lo anteriormente mencionado, en las Figuras 3.2, 3.3 y 3.4. Estas imágenes son correspondientes a la mamografía del caso del paciente 00001, mama izquierda con vista craneocaudal (P_00001_LEFT_CC), un caso del subconjunto de entrenamiento.

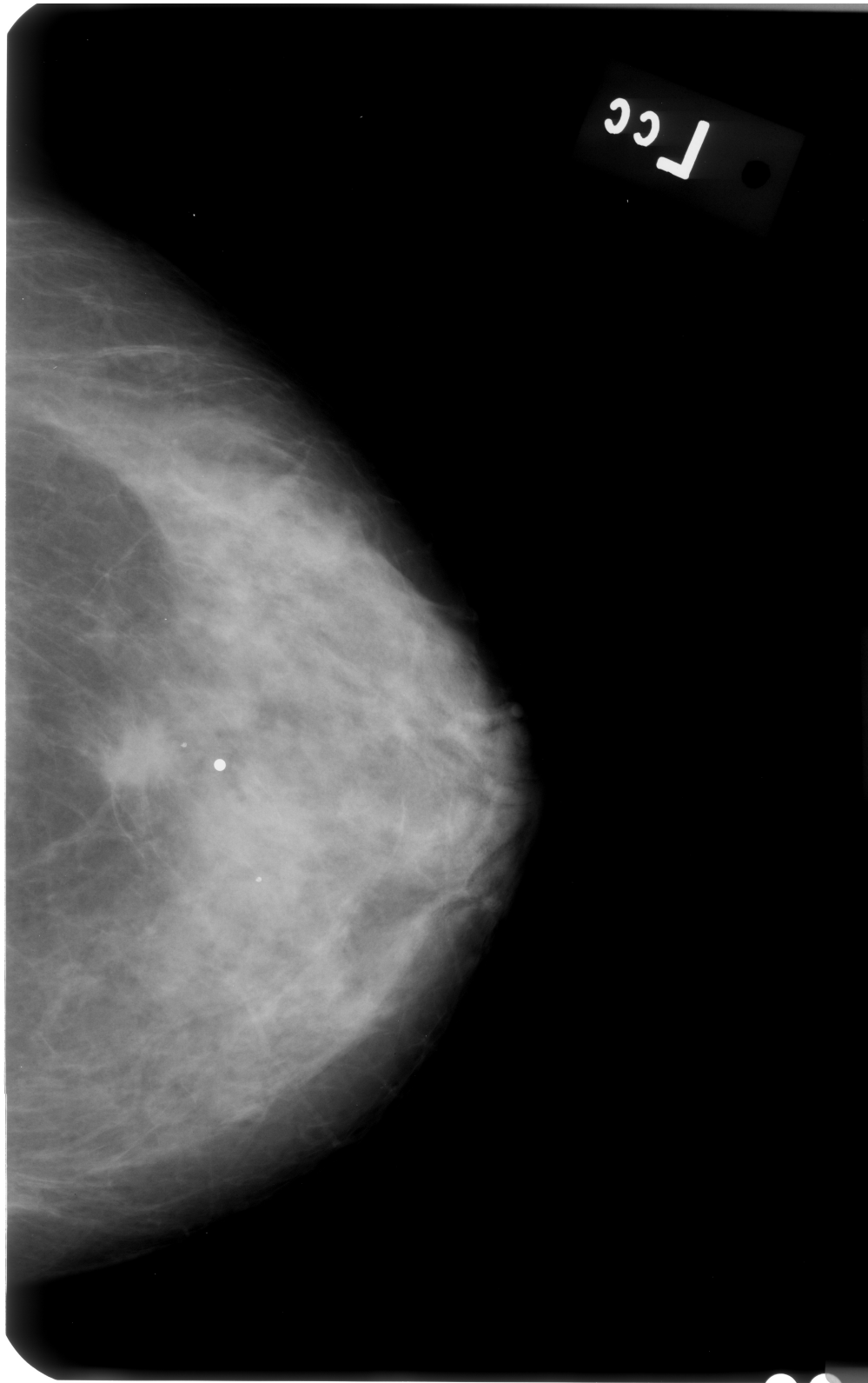


Figura 3.2: Mamografía completa

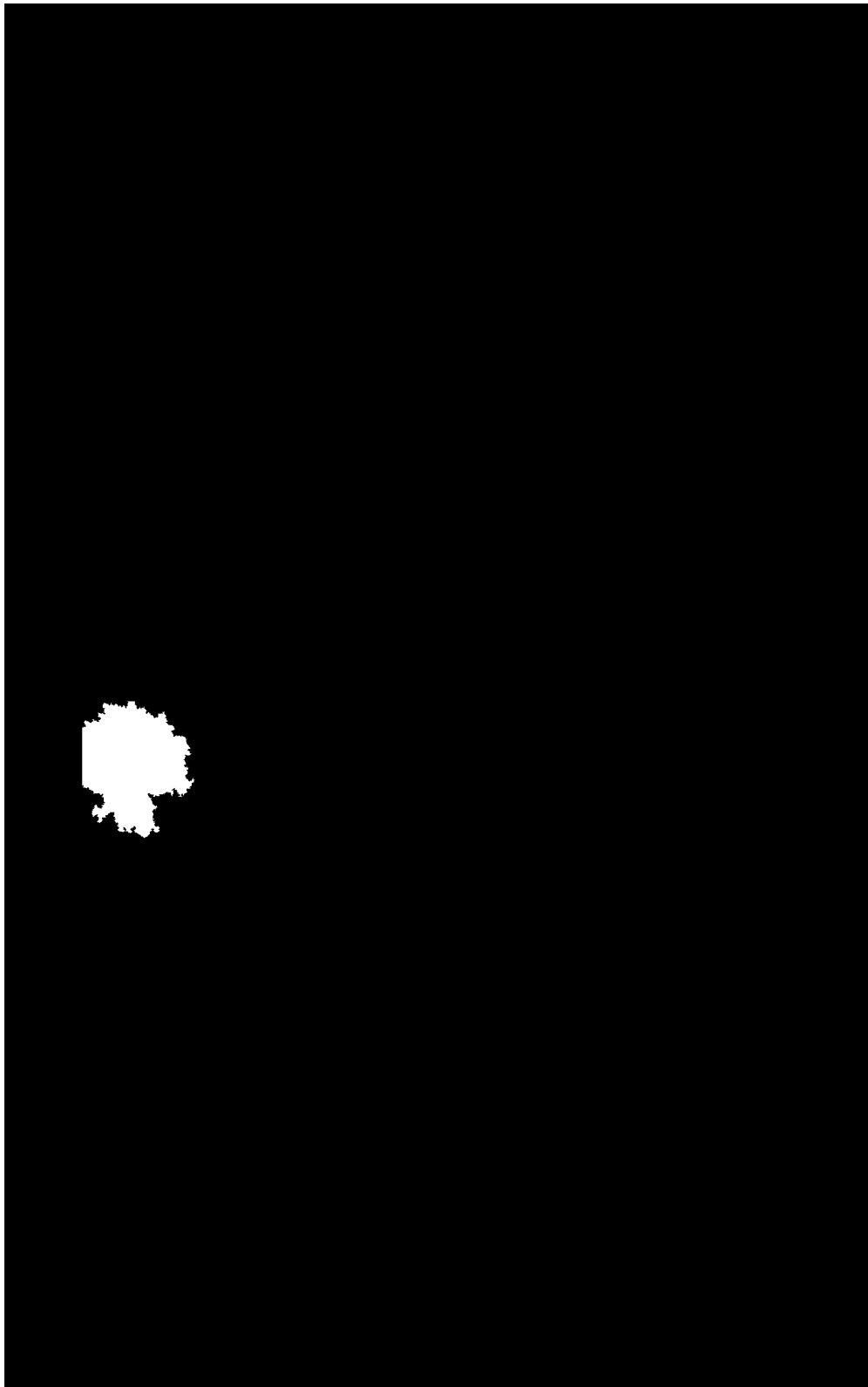


Figura 3.3: ROI Mask de la única masa presente en esta mamografía

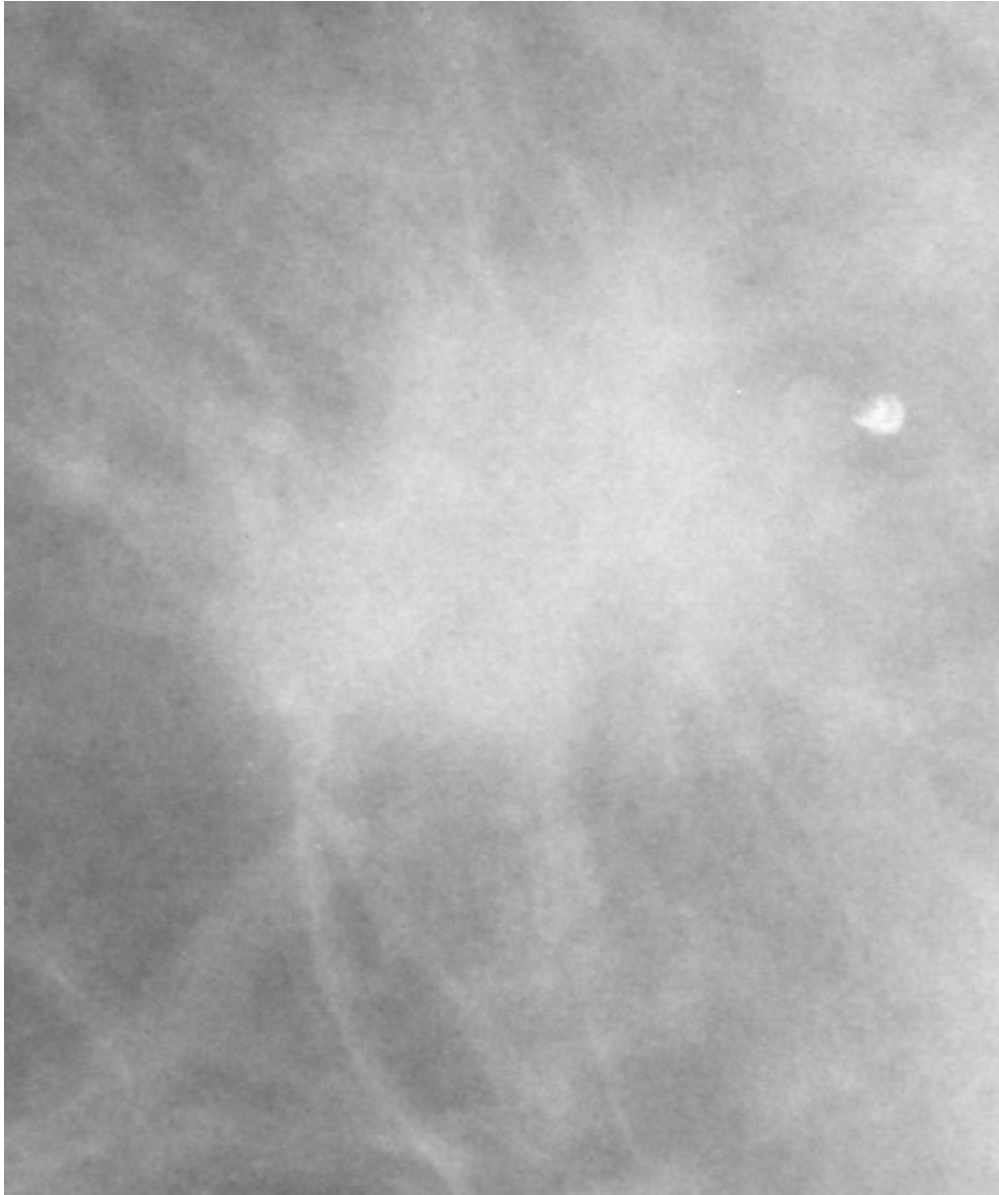


Figura 3.4: Imagen recortada de la única masa presente en esta mamografía

Una de las principales decisiones tomadas a partir de las imágenes fue la exclusión de las imágenes recortadas del trabajo a realizar, puesto que se determinó que no eran realmente necesarias dado que lo único necesario era la mamografía completa, para poder ubicar la o las masas, y las ROI masks para poder detectar la ubicación de cada masa. Así que a partir de aquí solo se realizó cargue, análisis y desarrolló con las mamografías completas y las ROI Mask.

En la Figura 3.5 se presentan histogramas de escala de grises para (a) la mamografía completa y (b) la ROI Mask. En la mamografía completa se observa una alta variación de intensidades, con predominio de píxeles cercanos a 0 (negro). En la ROI Mask, en cambio, únicamente aparecen valores 0 y 255, correspondientes a negro y blanco.

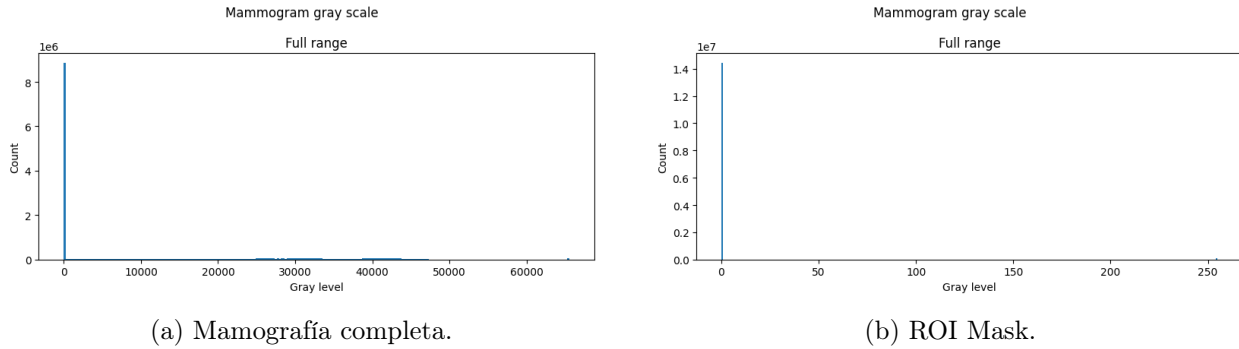


Figura 3.5: Histogramas de escala de grises para la mamografía completa y su ROI Mask.

3.1.6. Cálculo del centro de masa en las ROI Mask

Llegado a este punto la tarea a realizar era la de determinar las bounding boxes de las masas presentes en las mamografías del conjunto de datos para así poder extraer los datos que el modelo YOLO requiere para poder realizar el entrenamientos, entendiéndose como estos datos a la clase, el centro y el alto y el ancho de la bounding box que contiene la masa. Para esta tarea se decidió utilizar “la teoría de centro de masa”.

El centro de masa es el punto en el que se puede considerar concentrada toda la masa de un sistema sin alterar su comportamiento traslacional. Para un sistema de n partículas con masas m_i situadas en posiciones \mathbf{r}_i , el centro de masa \mathbf{r}_{CM} se define como el promedio ponderado de las posiciones:

$$\mathbf{r}_{CM} = \frac{1}{M} \sum_{i=1}^n m_i \mathbf{r}_i, \quad M = \sum_{i=1}^n m_i. \quad (3.1)$$

En dos dimensiones, escribiendo $\mathbf{r}_i = (x_i, y_i)$, las coordenadas del centro de masa quedan:

$$x_{CM} = \frac{1}{M} \sum_{i=1}^n m_i x_i, \quad y_{CM} = \frac{1}{M} \sum_{i=1}^n m_i y_i. \quad (3.2)$$

Sea $B(x, y)$ una máscara binaria de la región de interés (ROI), donde $B(x, y) = 1$ si el píxel (x, y) pertenece a la masa y $B(x, y) = 0$ en caso contrario. Si denotamos por \mathcal{R} al conjunto de píxeles de la masa y por $N = |\mathcal{R}|$ el número total de píxeles de la ROI, las coordenadas del centro de masa (centroide) se calculan como:

$$x_c = \frac{1}{N} \sum_{(x,y) \in \mathcal{R}} x, \quad y_c = \frac{1}{N} \sum_{(x,y) \in \mathcal{R}} y. \quad (3.3)$$

En este trabajo, la teoría del centro de masa se empleó para obtener el punto central de cada masa a partir de su ROI mask, considerando cada píxel de la región como una contribución de igual peso.

3.1.7. Generación y ajuste de bounding boxes

Antes de pasar a generar los bounding box a partir del centro de masa se encontró un pequeño problema con la marcación de las ROI Mask para las masas, y es que resultó que en los bordes de las ROI Mask dibujadas había un “residuo” de píxeles blancos. Es decir, a simple vista (como se podrá evidenciar en la imagen de ejemplo de ROI Mask) pareciera que todos los píxeles blancos presentes en la imagen están unidos en una sola masa, pero no es así. A nivel de pixel por pixel, hay presencia de píxeles blancos ligeramente separados de la “masa principal”, suelen ser 1 o 2 píxeles separados, a distancias de 1-3 píxeles.

El problema que se presenta esta en la detección de centros de masas, puesto que estos “píxeles residuales” resultan ser detectados a nivel algorítmico como “masas”, y acababan generándose bounding box para cada uno de ellos también, además de que terminaba generándose una bounding box para la “masa principal” que no era del todo exacta. En la Figura 3.7 se pueden evidenciar la señalización de algunos de estos “píxeles residuales”.

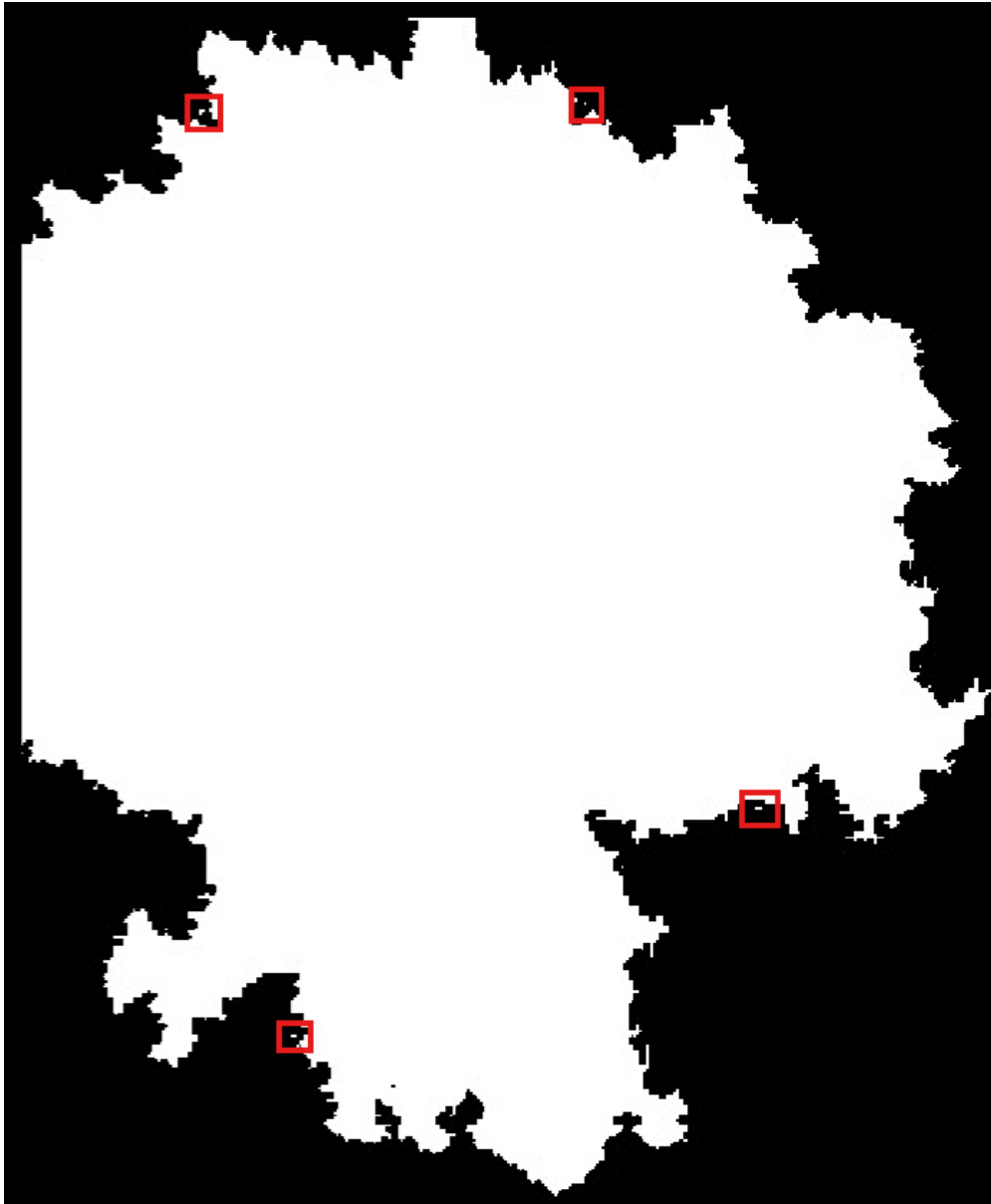


Figura 3.6: Ejemplo de ROI Mask con zoom realizado y con anotaciones

Para solucionar este problema se realizó una unión de los *pixeles residuales* con la *masa principal*.

Lo primero que se hace es detectar los componentes y clasificarlos como “masa grande” y “masas pequeñas”. Después se realiza una dilatación de la masa grande para detectar las “masas pequeñas” que entrarían a formar parte de la “masa principal”. Esto último se manejaría mediante un radio que delimita hasta donde llega la dilatación. Una vez se tenga detectada la masa pequeña, y este dentro del rango, se procede a crear un “puente” entre dicha masa y la masa grande para poder realizar una unión física. Este paso a paso se repite para cada masa pequeña que se encuentre dentro

del rango de dilatación de la masa grande. Y de esta forma finalmente se construye una sola ROI Mask con todas las masas pequeñas o “píxeles residuales” unidos.

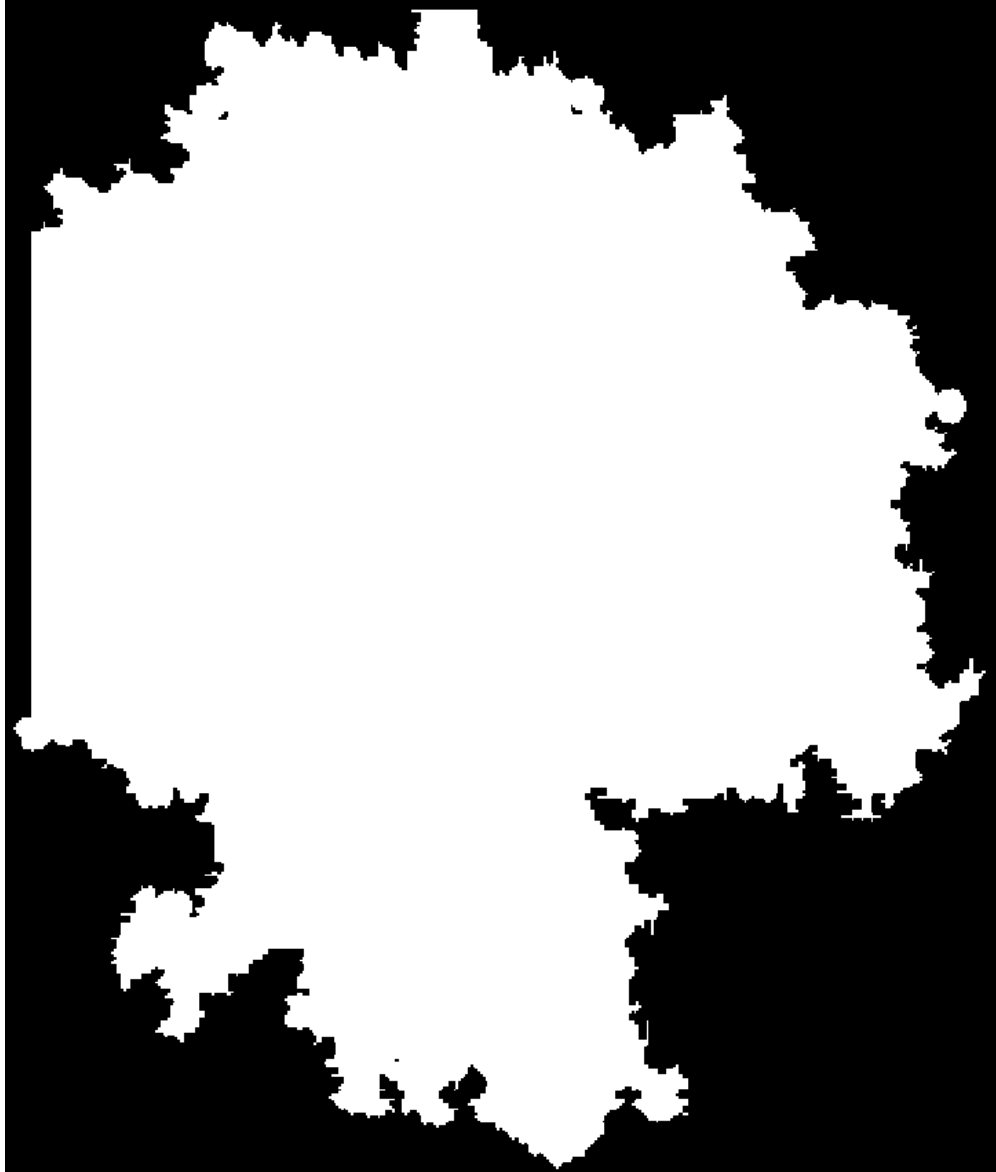


Figura 3.7: Ejemplo de ROI Mask con zoom y sin píxeles residuales

Teniendo una sola masa y su único centro de masa, sin “píxeles residuales”, se puede proceder a generar el bounding box. Cabe destacar que para mayor abarcamiento del proceso, este se diseñó para que sea capaz de hallar n centros de masa y generar n bounding boxes para n masas presentes en cada ROI Mask, suponiendo que no fuera una ROI Mask por masa sino una sola que contuviera todas las masas de la mamografía digital.

La generación de cada bounding box llegados a este punto simplemente consiste en delimitar la masa partiendo desde su centro de masas previamente calculado, al mismo tiempo que se calcula un porcentaje “extra” de margen de la bounding box con respecto a la masa (en este caso del 15%), esto con el objetivo de darle mayor “realismo” a la anotación. Hay que tener en cuenta que no es del todo realista que una bounding box delimite de manera perfecta con la masa, debe haber un ligero margen de error. Este margen se calcula de manera simétrica. En las figuras 3.9 y 3.10 se puede evidenciar como quedaría finalmente la ROI Mask y la mamografía completa con la bounding box calculada y anotada.

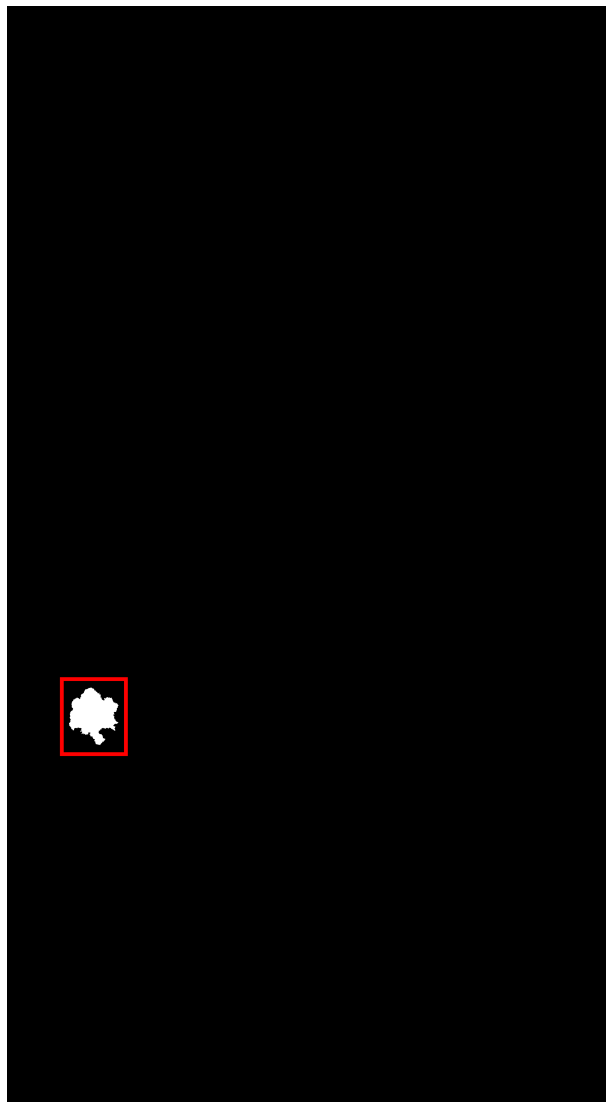


Figura 3.8: Ejemplo de ROI Mask con *bounding box* anotada.

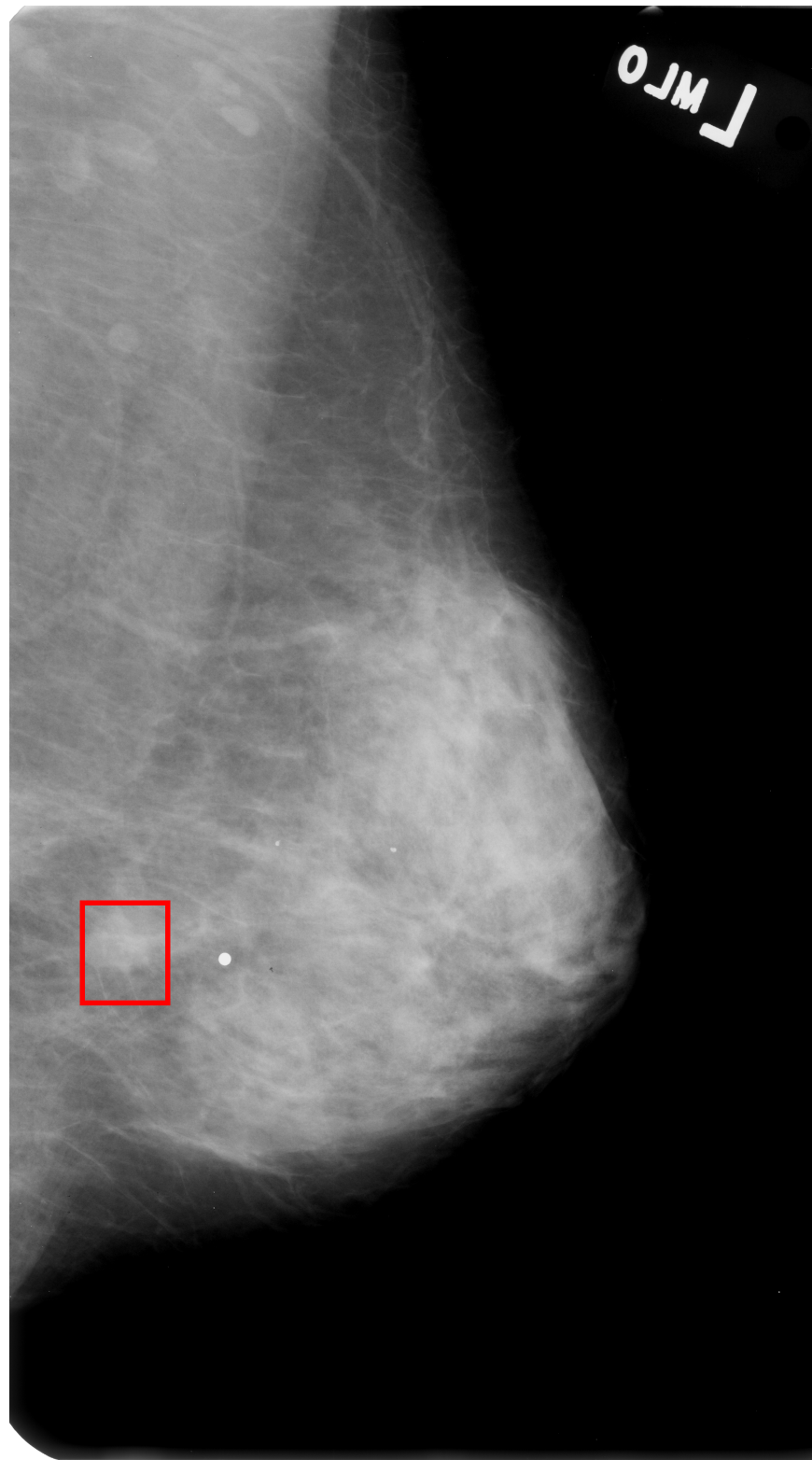


Figura 3.9: Ejemplo de mamografía completa con *bounding box* anotada.

3.1.8. Generación de la estructura de carpetas para YOLO

Una vez generadas y validadas las *bounding boxes* asociadas a cada masa, fue necesario adaptar el conjunto de datos al formato de entrada esperado por el framework YOLO. Para ello se diseñó y construyó una estructura de directorios específica, que permite separar claramente las imágenes y sus anotaciones en los subconjuntos de entrenamiento, validación y prueba, y facilita la integración con los archivos de configuración del modelo.

En primer lugar, se definió un directorio raíz del proyecto, dentro del cual se crearon las carpetas `images` y `labels`. Cada una de estas carpetas se subdividió, a su vez, en tres subdirectorios correspondientes a los subconjuntos de datos empleados durante el entrenamiento del modelo: `train`, `val` y `test`. La estructura final puede resumirse de la siguiente manera:

```
dataset/  
  images/  
    train/  
    val/  
    test/  
  labels/  
    train/  
    val/  
    test/
```

Las imágenes mamográficas con presencia de masas se distribuyeron en estos subconjuntos de forma controlada, procurando evitar que un mismo paciente apareciera simultáneamente en más de un subconjunto, con el fin de prevenir fugas de información (*data leakage*) entre entrenamiento, validación y prueba. Para cada imagen ubicada en `images/train`, `images/val` o `images/test`, se reservó un archivo de etiqueta con el mismo nombre base y extensión `.txt` en el subdirectorio correspondiente de `labels`.

La generación de esta estructura se realizó de manera automática mediante scripts, a partir de las rutas de archivo originales proporcionadas por CBIS-DDSM. De este modo, se garantizó la trazabilidad entre los metadatos, las imágenes originales, las máscaras de ROI y las nuevas anotaciones en formato YOLO, dejando el conjunto de datos preparado para las etapas de creación de etiquetas y entrenamiento del modelo descritas en las siguientes subsecciones.

3.1.9. Creación de etiquetas en formato YOLO

Una vez definida la estructura de carpetas para el conjunto de datos, el siguiente paso consistió en generar las anotaciones en el formato requerido por YOLO. Para ello se implementó una función que integra los procedimientos descritos en las secciones previas: el cargue de cada imagen, el cálculo del centro de masa a partir de la ROI mask y la generación de las *bounding boxes* asociadas a cada lesión.

En el formato YOLO, cada objeto presente en una imagen se representa en un archivo de texto mediante una línea con la siguiente estructura:

$$\text{clase } x_c \ y_c \ w \ h, \quad (3.4)$$

donde x_c e y_c son las coordenadas normalizadas del centro de la *bounding box*, y w y h corresponden al ancho y alto normalizados de la caja, respectivamente. Todas estas cantidades se expresan en el rango $[0, 1]$, dividiendo por el ancho W y el alto H de la imagen original. En este trabajo se definió una única clase (0) correspondiente a la presencia de masa mamaria, de modo que cada línea del archivo de anotaciones tiene la forma:

$$0 \ x_c \ y_c \ w \ h. \quad (3.5)$$

Para cada mamografía en el conjunto de datos se realizan los siguientes pasos:

1. Recorre las mascararas de región de interés (ROI) asociadas a esa mamografía. Para cada una de ellas se carga la imagen de manera local mientras se trabaja con ella.
2. A partir de la ROI mask se calcula el centro de masa de la región (*mass center*) empleando la técnica previamente mencionada.
3. Con los centros de masa obtenidos se generan las *bounding boxes* mínimas que enmarcan cada masa. Estas cajas se definen en coordenadas de píxel sobre la imagen original.
4. Para cada *bounding box* se calculan las coordenadas normalizadas en el formato YOLO. Sea una caja delimitada por las coordenadas de píxel $(x_{\text{mín}}, y_{\text{mín}})$ y $(x_{\text{máx}}, y_{\text{máx}})$, y sea $W \times H$ el tamaño de la imagen. Entonces:

$$x_c = \frac{x_{\text{mín}} + x_{\text{máx}}}{2W}, \quad y_c = \frac{y_{\text{mín}} + y_{\text{máx}}}{2H}, \quad w = \frac{x_{\text{máx}} - x_{\text{mín}}}{W}, \quad h = \frac{y_{\text{máx}} - y_{\text{mín}}}{H}. \quad (3.6)$$

5. Con estos valores se construye una línea de texto por cada masa detectada, de la forma “0 x_c y_c w h ”, que se almacena en una lista de etiquetas.
6. Una vez procesadas todas las máscaras asociadas a la imagen, se escriben dichas líneas en un archivo `.txt` ubicado en el directorio `labels/mode` (siendo `mode` los valores `train`, `val` o `test`), utilizando como nombre base el identificador de la imagen.
7. Finalmente, se carga la imagen mamográfica correspondiente, se convierte al formato PNG y se guarda en el directorio `images/mode` con el mismo nombre base que el archivo de etiquetas.

De esta manera, para cada imagen incluida en los subconjuntos de entrenamiento, validación o prueba, se genera de forma automática un archivo de anotaciones en formato YOLO que contiene las coordenadas normalizadas de todas las masas presentes. Este procedimiento garantiza la consistencia entre las imágenes, sus *bounding boxes* y las clases asignadas, y deja el conjunto de datos listo para ser utilizado directamente en el entrenamiento del modelo YOLO.

3.2. Entrenamiento

3.2.1. Entrenamiento del modelo YOLO

Una vez construida la estructura de carpetas del conjunto de datos (`train`, `val` y `test`) y generados los archivos de anotaciones en formato YOLO, se lleva a cabo el entrenamiento del modelo de detección de masas. Para ello se utiliza la implementación de la familia YOLO disponible en la librería *Ultralytics*, que permite definir el problema de detección mediante un archivo de configuración `.yaml`. En dicho archivo se indican las rutas de las imágenes de cada partición, la ubicación de las etiquetas y las clases presentes en el problema (en este trabajo, una única clase denominada `mass`).

El proceso de entrenamiento se organiza en épocas (*epochs*), en las que el modelo recorre todas las imágenes del conjunto de entrenamiento en pequeños grupos (*batches*). Al final de cada época se evalúa el modelo sobre el conjunto de validación, calculando métricas estándar de detección de objetos como la precisión (P), el *recall* (R) y el promedio de precisión media (*mean Average Precision*, mAP) para distintos niveles de solapamiento entre las cajas predichas y las anotaciones reales. Estas métricas permiten seguir la evolución del entrenamiento, seleccionar automáticamente el *checkpoint* con mejor rendimiento y detener el proceso cuando ya no se observan mejoras significativas (*early stopping*).

En este trabajo, el conjunto de prueba no incluye mamografías negativas: todas las imágenes contienen al menos una masa anotada. Por construcción, una imagen sin detecciones implica necesariamente la presencia de falsos negativos a nivel de lesión. En consecuencia, además de maximizar métricas globales como mAP, se calibró el punto de operación del detector, definido por el umbral de confianza *conf* y el umbral de solapamiento *IoU* utilizado en la *Non-Maximum Suppression* (NMS). Estos parámetros se ajustaron para priorizar la sensibilidad del sistema y reducir la probabilidad de omitir masas.

3.2.2. Modelo base: YOLOv8m

Para el desarrollo de este trabajo se seleccionó como modelo base **YOLOv8m** (*You Only Look Once*, versión 8), perteneciente a la familia de modelos de detección de objetos de una sola etapa (*one-stage detectors*). En este tipo de enfoques, la detección se realiza en un único paso: a partir de una imagen de entrada, el modelo predice directamente un conjunto de *bounding boxes* y una puntuación de confianza asociada a cada predicción, sin necesidad de una etapa adicional de propuesta de regiones. Esto permite un flujo de inferencia eficiente y, al mismo tiempo, obtener predicciones de localización y clasificación de manera conjunta.

¿Por qué YOLOv8m? Dentro de la familia YOLOv8 existen variantes de distinto tamaño (por ejemplo, n , s , m , l , x), las cuales ofrecen un compromiso entre costo computacional y capacidad de representación. La variante **YOLOv8m** fue seleccionada por ser un punto intermedio adecuado para este problema: posee suficiente capacidad para modelar patrones complejos presentes en

mamografías (diferencias sutiles de contraste y variaciones de forma de las masas), sin llegar a los requerimientos computacionales elevados de modelos más grandes.

Adicionalmente, el conjunto CBIS-DDSM utilizado en este trabajo tiene un tamaño moderado (del orden de miles de imágenes), lo cual favorece el uso de un modelo con capacidad intermedia. Modelos demasiado pequeños pueden subajustar (*underfitting*) y perder sensibilidad en masas pequeñas, mientras que modelos excesivamente grandes pueden incrementar el riesgo de sobreajuste (*overfitting*) y demandar mayor tiempo de entrenamiento sin un beneficio proporcional. Por estas razones, YOLOv8m ofrece un balance apropiado para un proyecto de pregrado.

Proceso de entrenamiento a partir del modelo base. El entrenamiento se realiza ajustando los parámetros del modelo mediante retropropagación (*backpropagation*). En cada iteración, se procesa un lote de imágenes (*batch*) y se comparan las predicciones con las anotaciones reales. La pérdida total de detección combina típicamente: (i) una componente de localización, que penaliza diferencias entre las cajas predichas y las cajas reales; y (ii) una componente de confianza / clasificación, que incentiva predicciones con alta confianza cuando existe una masa y baja confianza cuando no corresponde a una anotación.

En la implementación de *Ultralytics*, el entrenamiento se configura a través de un archivo `.yaml` del conjunto de datos y parámetros como el tamaño de entrada (*imgsz*), el número de épocas y el tamaño del lote. Al final de cada época se evalúa el desempeño en el conjunto de validación, registrando métricas como precisión, *recall* y mAP, lo que permite seleccionar automáticamente el mejor *checkpoint*. Finalmente, el modelo seleccionado se evalúa en el conjunto de prueba y se ajusta el punto de operación (umbral de confianza y parámetros de NMS) para priorizar sensibilidad, dado el contexto del problema.

3.2.3. Entrenamiento del modelo y ajuste de hiperparámetros

El entrenamiento del detector se realizó utilizando la implementación de *Ultralytics* para la familia YOLOv8, partiendo del modelo pre-entrenado `yolov8m.pt`. Este archivo corresponde a la variante *medium* (“m”) de YOLOv8 para detección, cuyos pesos iniciales fueron entrenados sobre el conjunto *COCO* (80 clases) y se recomienda como punto de partida para *transfer learning* en conjuntos de datos personalizados [36, 37, 38].

Modelo pre-entrenado (pesos base). YOLOv8m presenta un balance intermedio entre capacidad y costo computacional. De acuerdo con la documentación oficial de *Ultralytics* para detección en *COCO*, **YOLOv8m** (con tamaño de entrada 640) cuenta con aproximadamente **25.9 millones de parámetros** y requiere alrededor de **78.9 GFLOPs** por inferencia, alcanzando un rendimiento reportado de **50.2 mAP@0.5:0.95** en validación sobre *COCO* [36]. Estos valores justifican su selección como base para el problema de detección de masas, dado que ofrece capacidad suficiente para capturar patrones complejos sin los costos elevados de variantes mayores (p. ej., `yolov8l` o `yolov8x`).

Hiperparámetros principales. Los hiperparámetros ajustados explícitamente (vía CLI) fueron:

- **Épocas** (`epochs=100`): número máximo de pasadas completas por el conjunto de entrenamiento. Ultralytics permite *early stopping* mediante el parámetro `patience`; en ausencia de especificación, se mantiene el valor por defecto (`patience=100`) [39].
- **Tamaño de entrada** (`imgsz=640`): las imágenes se redimensionan a una entrada cuadrada de 640×640 .
- **Tamaño de lote** (`batch=8`): número de imágenes procesadas antes de una actualización de pesos. Este valor suele estar condicionado por la memoria disponible en GPU; lotes mayores estabilizan el gradiente pero requieren más memoria [39].

Hiperparámetros por defecto relevantes. Además de los parámetros anteriores, la librería define valores por defecto que afectan el entrenamiento. Por ejemplo, `optimizer=auto`, `lr0=0.01`, `momentum=0.937`, `weight_decay=0.0005`, `warmup_epochs=3.0`, entre otros [39]. En este trabajo, estos valores se mantuvieron sin modificación en el entrenamiento base, con el fin de priorizar una configuración reproducible y ampliamente utilizada en la documentación oficial.

Evaluación durante entrenamiento (validación por época). Al finalizar cada época, Ultralytics evalúa el modelo sobre el conjunto de validación y reporta métricas estándar de detección como precisión (P), *recall* (R) y $\text{mAP}@0.5$ / $\text{mAP}@0.5:0.95$. Estas métricas permiten seleccionar automáticamente el *checkpoint* con mejor desempeño y registrar la evolución del aprendizaje a lo largo de las épocas [36, 37].

3.2.4. Evidencia visual del entrenamiento

Con el fin de documentar el proceso de aprendizaje y complementar las métricas numéricas, se incorporan las figuras generadas automáticamente por Ultralytics durante el entrenamiento. Estas salidas permiten observar (i) la evolución de las pérdidas y métricas por época, (ii) el comportamiento del modelo frente a distintos umbrales de confianza, y (iii) ejemplos cualitativos de predicciones sobre el conjunto de validación.

Evolución por época. La Figura 3.10 presenta la evolución de las métricas de desempeño a lo largo de las épocas de entrenamiento. En esta figura se espera observar una tendencia decreciente de las pérdidas y una estabilización o incremento gradual de métricas como mAP , lo cual sugiere convergencia del modelo.

Curvas PR y curvas por umbral de confianza. Para analizar el compromiso entre precisión y sensibilidad, se presentan curvas que describen el desempeño del detector ante distintos umbrales de confianza. La curva Precisión–Recall (PR) resume el desempeño global para diferentes puntos de operación, mientras que las curvas por umbral permiten identificar valores de `conf` que priorizan sensibilidad o reducen falsos positivos.

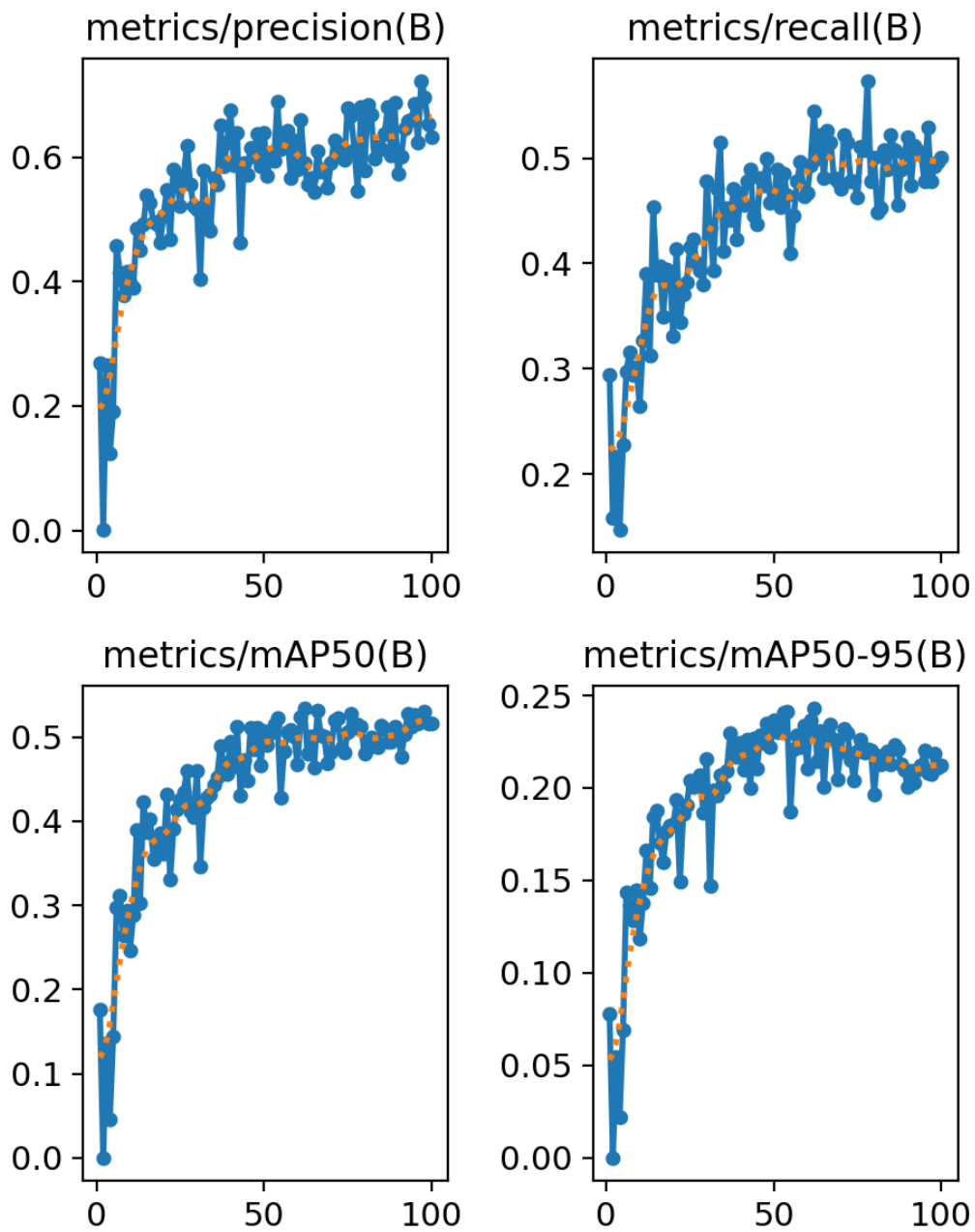


Figura 3.10: Evolución de métricas del entrenamiento por época.

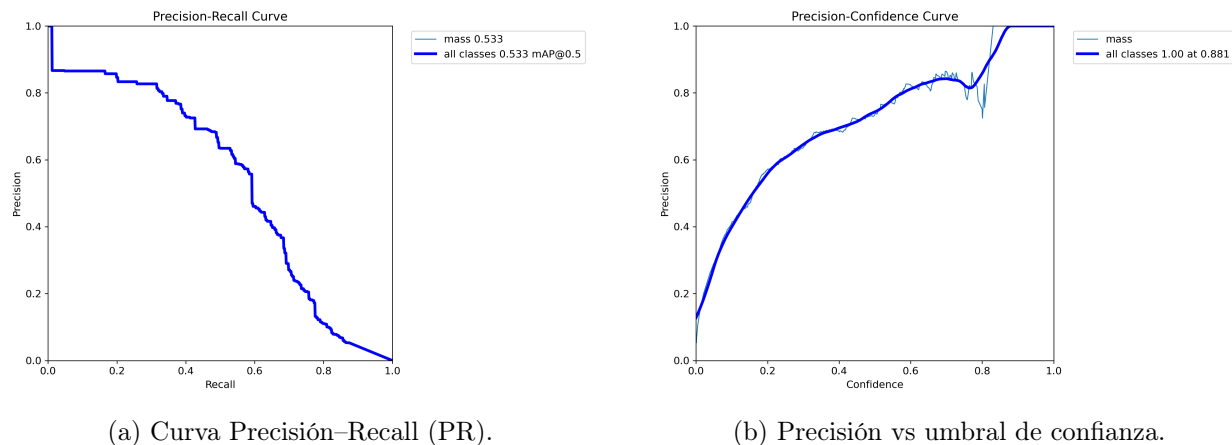


Figura 3.11: Curvas de desempeño.

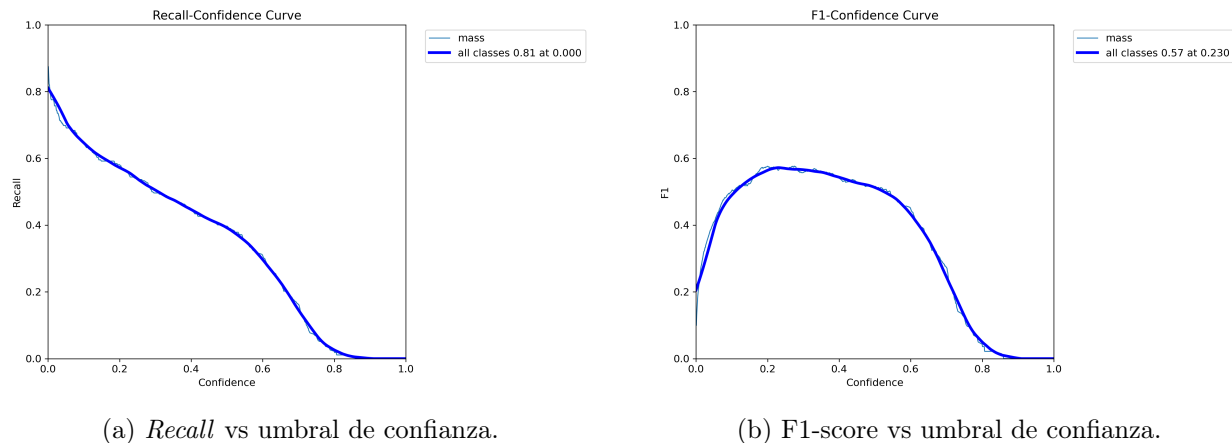


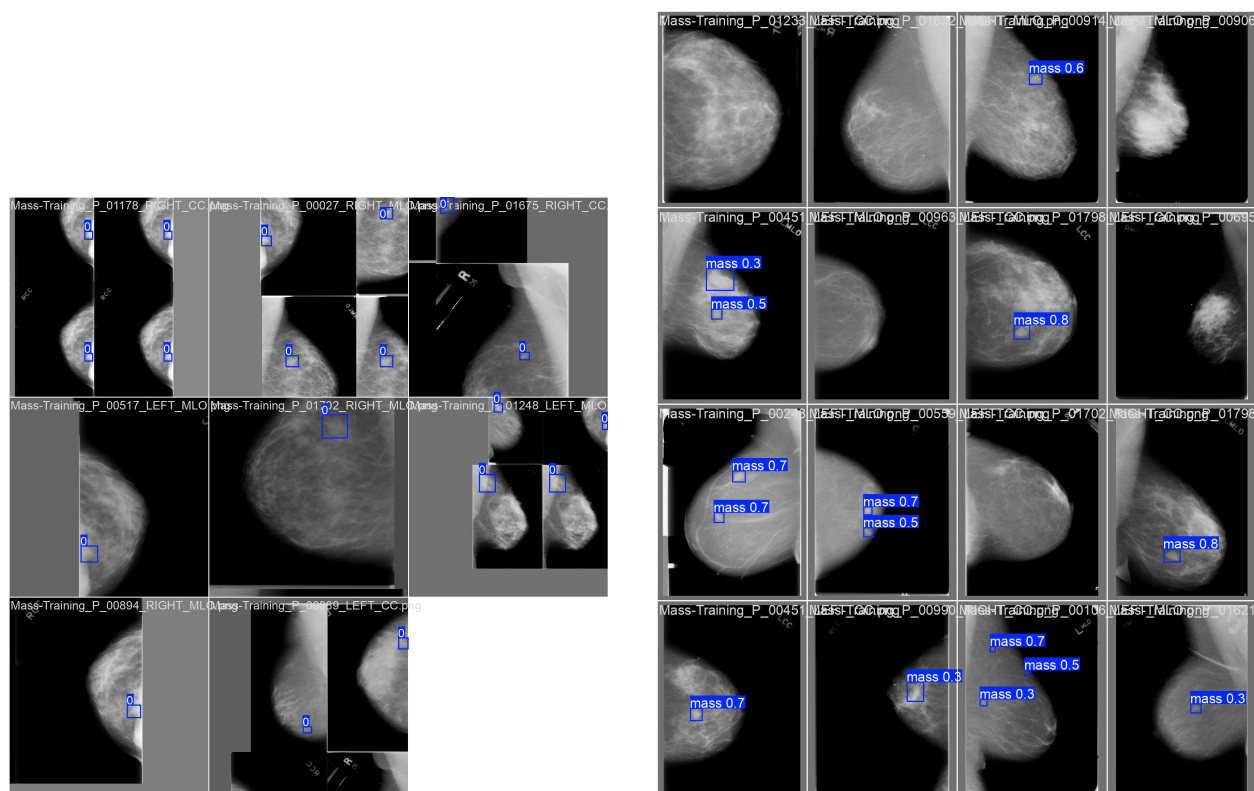
Figura 3.12: Curvas de desempeño.

Ejemplos cualitativos (lotes y predicciones). La Figura 3.13 muestra ejemplos de lotes de entrenamiento y predicciones sobre validación. Estas imágenes permiten inspeccionar cualitativamente el ajuste de las *bounding boxes* y complementar las métricas cuantitativas.

Métricas finales en validación. El Cuadro 3.2 resume las métricas obtenidas en el conjunto de validación para el *checkpoint* seleccionado como mejor modelo (*best.pt*).

Cuadro 3.2: Métricas del mejor *checkpoint* en el conjunto de validación.

Conjunto	P	R	mAP@0,5	mAP@0,5 : 0,95
Validación	0.586	0.544	0.533	0.243



(a) Ejemplo de lote de entrenamiento (mosaico).

(b) Predicciones sobre un lote de validación.

Figura 3.13: Ejemplos visuales generados durante el entrenamiento.

Resultados

4.1. Selección final del modelo y evaluación clínica

Una vez completado el entrenamiento base con **YOLOv8m** a resolución `imgsz=640`, se realizaron pruebas adicionales con el objetivo de mejorar el desempeño del detector. Entre estas pruebas se incluye un *fine-tuning* a mayor resolución (`imgsz=1536`) y una variante de inferencia con `augment=True`. Sin embargo, los resultados obtenidos no evidenciaron una mejora consistente en el conjunto de validación/prueba; en algunos casos incluso se observó degradación del desempeño. Por esta razón, se seleccionó como modelo final el detector YOLOv8m descrito en la sección anterior, al representar el mejor compromiso entre rendimiento, estabilidad y costo computacional.

4.1.1. Motivación de la evaluación adicional

Si bien Ultralytics reporta métricas estándar de detección (precisión, *recall* y mAP) sobre el conjunto de prueba, en este trabajo se buscó complementar estas medidas con indicadores más interpretables desde un punto de vista clínico. En particular, al tratarse de mamografías con al menos una masa en cada imagen, resulta de interés cuantificar no solo si el modelo detecta masas, sino también *qué tan bien localiza* dichas regiones y cuántas falsas alarmas produce.

Por lo anterior, la evaluación final se basó en dos métricas principales:

- **Índice de Jaccard (IoU):** mide el solapamiento entre la región predicha y la anotación real (Ground Truth). Un valor alto indica una localización precisa y una alta coincidencia espacial entre ambas regiones.
- **FROC (Sensitivity vs FPPI):** evalúa la sensibilidad a nivel de lesión frente a falsos positivos por imagen (*False Positives Per Image*, FPPI). Esta métrica permite analizar el compromiso entre detectar la mayor cantidad de masas posibles y limitar el número de falsas alarmas por imagen.

4.1.2. Pipeline de inferencia y post-procesamiento

Para calcular de forma consistente el Índice de Jaccard y construir el análisis FROC, se implementó un pipeline de inferencia y post-procesamiento sobre el conjunto de prueba. Este pipeline es necesario porque, en múltiples ocasiones, el modelo puede producir varias cajas altamente superpuestas sobre una misma región de interés, lo cual dificulta el cálculo de solapamiento (IoU) y puede inflar artificialmente el número de falsos positivos.

El flujo general se resume en la Figura 4.1 y puede describirse en los siguientes pasos:

1. **Inferencia YOLOv8m:** se generan las predicciones iniciales del modelo sobre cada imagen del conjunto de prueba.
2. **NMS interno:** durante la inferencia, el modelo aplica una etapa estándar de *Non-Maximum Suppression* para eliminar predicciones redundantes.
3. **Extra NMS (post-procesamiento):** se aplica una segunda etapa de NMS sobre las predicciones resultantes (Extra NMS), con el objetivo específico de eliminar cajas duplicadas altamente superpuestas dentro de una misma región.
4. **Cálculo de métricas:** con las predicciones finales se calcula (i) el Índice de Jaccard entre GT y predicción, y (ii) la sensibilidad y FPPI para construir la curva FROC (variando el umbral de confianza).

Diagrama del pipeline. La Figura 4.1 presenta un diagrama de bloques que resume el flujo general de evaluación implementado en este trabajo: *Imagen* \rightarrow *Predicción YOLO* \rightarrow *Extra NMS* \rightarrow *cálculo de métricas (Jaccard/FROC)*. Este diagrama facilita la comprensión del procedimiento seguido y sirve como puente hacia la presentación de los resultados obtenidos.

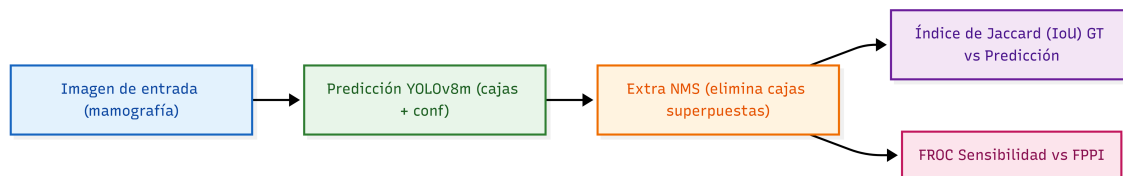


Figura 4.1: Diagrama del pipeline de inferencia y post-procesamiento para la evaluación final (Extra NMS + cálculo de métricas).

4.1.3. Evidencia visual del post-procesamiento (Extra NMS)

Con el fin de validar que el post-procesamiento elimina predicciones redundantes sin eliminar detecciones relevantes, se incluyen ejemplos comparativos de predicciones *antes* y *después* de aplicar Extra NMS. En la Figura 4.2 se muestran casos representativos donde se observa la reducción de cajas superpuestas en una misma región, manteniendo la caja final más representativa.

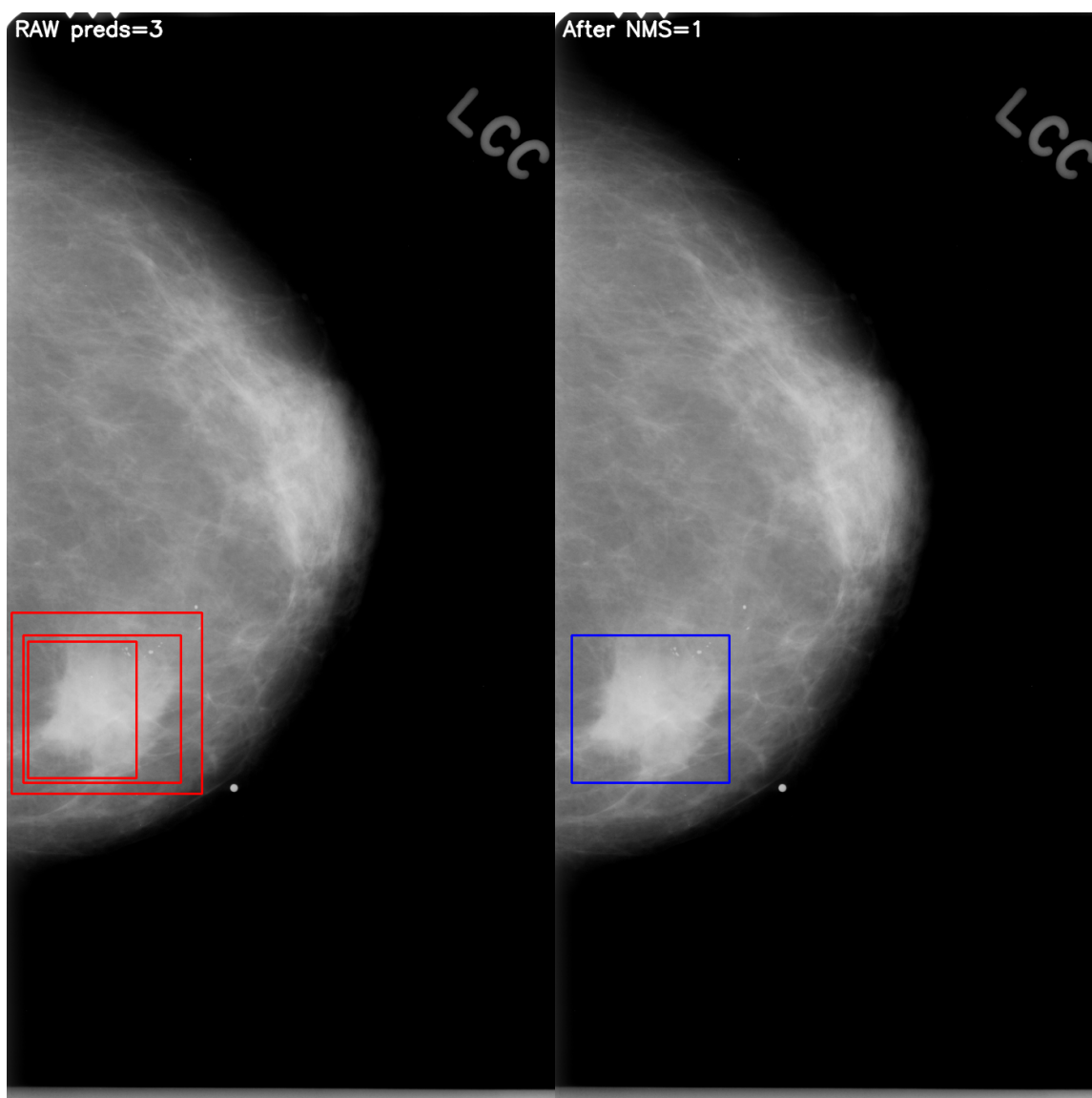


Figura 4.2: Ejemplo de predicción antes y después de aplicar Extra NMS. La etapa Extra NMS reduce cajas redundantes altamente superpuestas dentro de una misma región.

4.1.4. Resultados con Índice de Jaccard

Índice de Jaccard en presencia de múltiples masas. Dado que una misma mamografía puede contener más de una masa anotada, el cálculo del Índice de Jaccard (IoU) requiere un procedimiento de emparejamiento entre *bounding boxes* predichas y anotaciones reales. En este trabajo, las predicciones finales (posteriores a *Extra NMS*) se emparejan con las anotaciones mediante un criterio uno-a-uno (*one-to-one matching*), asignando cada predicción al *ground truth* con el mayor

IoU, siempre que dicho solapamiento supere un umbral ($\text{IoU} \geq 0,5$); de lo contrario, la predicción se considera un falso positivo.

Para resumir el desempeño por imagen, se calcula un valor J_{img} promediando el IoU de cada masa real con su predicción emparejada. En el caso de masas no detectadas (falsos negativos), se asigna $\text{IoU} = 0$ para reflejar la ausencia de localización. De este modo, el índice por imagen se define como:

$$J_{\text{img}} = \frac{1}{m} \sum_{i=1}^m \text{IoU}(g_i, p_{\pi(i)}),$$

donde m es el número de masas anotadas en la imagen, g_i representa cada anotación y $p_{\pi(i)}$ la predicción emparejada; si una masa no tiene predicción asignada, su contribución al promedio es cero.

A partir de las predicciones finales (posteriores al post-procesamiento con *Extra NMS*), se calculó el Índice de Jaccard (IoU) para cada imagen del conjunto de prueba, comparando la(s) predicción(es) con su correspondiente *ground truth*. El Cuadro 4.1 presenta la distribución de imágenes por rangos de Jaccard.

Cuadro 4.1: Distribución de imágenes del conjunto de prueba por rangos del Índice de Jaccard (IoU).

Rango Jaccard	Cantidad de imágenes	Porcentaje
$\geq 0,80$	49	13.57 %
0,80–0,60	94	26.04 %
0,60–0,40	33	9.14 %
0,40–0,20	26	7.20 %
$< 0,20$	159	44.04 %
Total	361	100.00 %

En particular, se observa que 143 de 361 imágenes presentan un solapamiento alto (Jaccard $\geq 0,60$), lo que corresponde al **39.61 %** del conjunto de prueba. Este resultado sugiere que, en una fracción considerable de casos, el modelo no solo detecta la presencia de la lesión, sino que también localiza la región con buena concordancia espacial con la anotación.

Evidencia visual. Para respaldar los resultados obtenidos con el Índice de Jaccard, se incluyen grillas de ejemplos cualitativos agrupados por rangos: (i) casos de Jaccard alto ($\geq 0,80$) y (ii) casos de Jaccard bajo ($< 0,40$). En cada grilla se utiliza una visualización tipo *Jaccard overlay* (GT vs predicción), la cual permite evidenciar de forma directa el grado de solapamiento entre la región anotada y la región detectada. Las Figuras 4.3–4.4 muestran ejemplos representativos de estos escenarios.

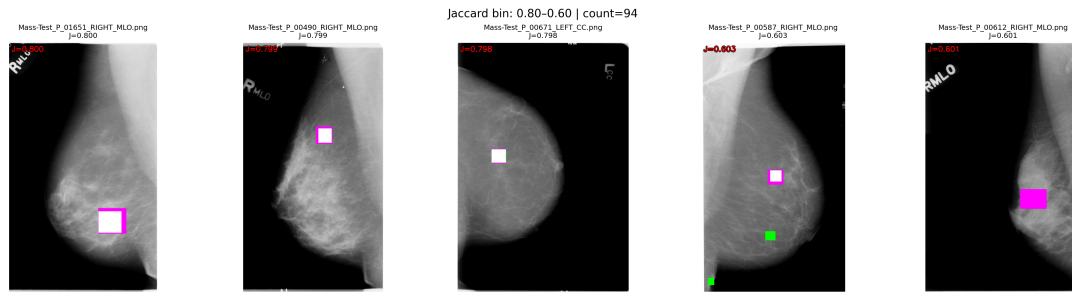


Figura 4.3: Ejemplos de alto solapamiento (Jaccard alto) usando visualización tipo *overlay* (GT vs predicción).

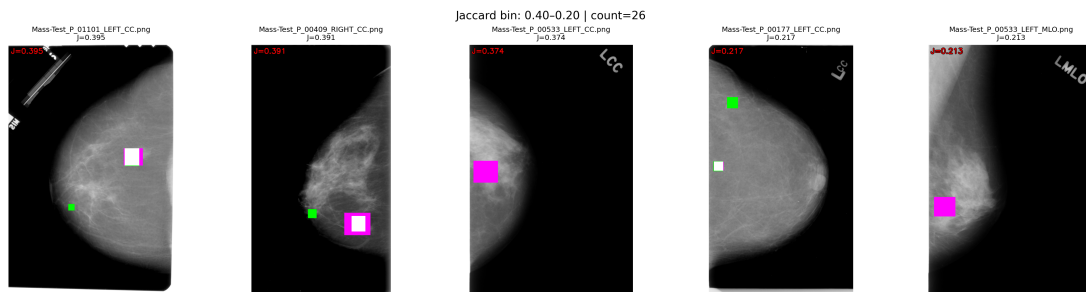


Figura 4.4: Ejemplos de bajo solapamiento (Jaccard bajo) usando visualización tipo *overlay* (GT vs predicción).

4.1.5. Resultados con FROC (Sensitivity vs FPPI)

Con el objetivo de evaluar el detector bajo un criterio más cercano a un entorno de apoyo clínico, se construyó una curva FROC, la cual relaciona la sensibilidad a nivel de lesión con el número de falsos positivos por imagen (*False Positives Per Image*, FPPI). Para la correspondencia entre predicciones y anotaciones se utilizó un criterio de *matching* uno-a-uno con umbral de IoU $\geq 0,5$, y se aplicó *Extra NMS* con umbral IoU = 0,5 para eliminar predicciones duplicadas altamente superpuestas.

El Cuadro 4.2 resume la sensibilidad máxima alcanzada bajo distintos límites de FPPI, junto con el umbral de confianza (*conf*) correspondiente al mejor punto bajo cada restricción.

Cuadro 4.2: Resumen FROC: sensibilidad máxima alcanzada bajo distintos límites de FPPI (IoU *matching* $\geq 0,5$, *Extra NMS* IoU = 0,5).

Límite FPPI	Sensibilidad	Umbral <i>conf.</i>	FPPI observado
0.5	0.5435	0.20	0.4903
1.0	0.6332	0.06	0.9889
2.0	0.6860	0.02	1.9862

Adicionalmente, el área bajo la curva FROC en el rango 0–2 FPPI fue de **AUC-FROC** = 1,1836. En términos prácticos, estos resultados evidencian el compromiso entre sensibilidad y falsas alarmas: al permitir un mayor FPPI (disminuyendo el umbral de confianza), se incrementa la sensibilidad, aunque con un aumento concomitante en el número de predicciones sin correspondencia.

Evidencia visual. La Figura 4.5 presenta la curva FROC completa, en la cual se marcan los puntos de operación asociados a los límites de FPPI reportados en el Cuadro 4.2 (FPPI \leq 0,5, 1,0 y 2,0). Esta visualización permite apreciar de manera global el compromiso entre sensibilidad y falsas alarmas por imagen, así como sustentar la selección del umbral de confianza (**conf**) utilizado para la evaluación final del modelo.

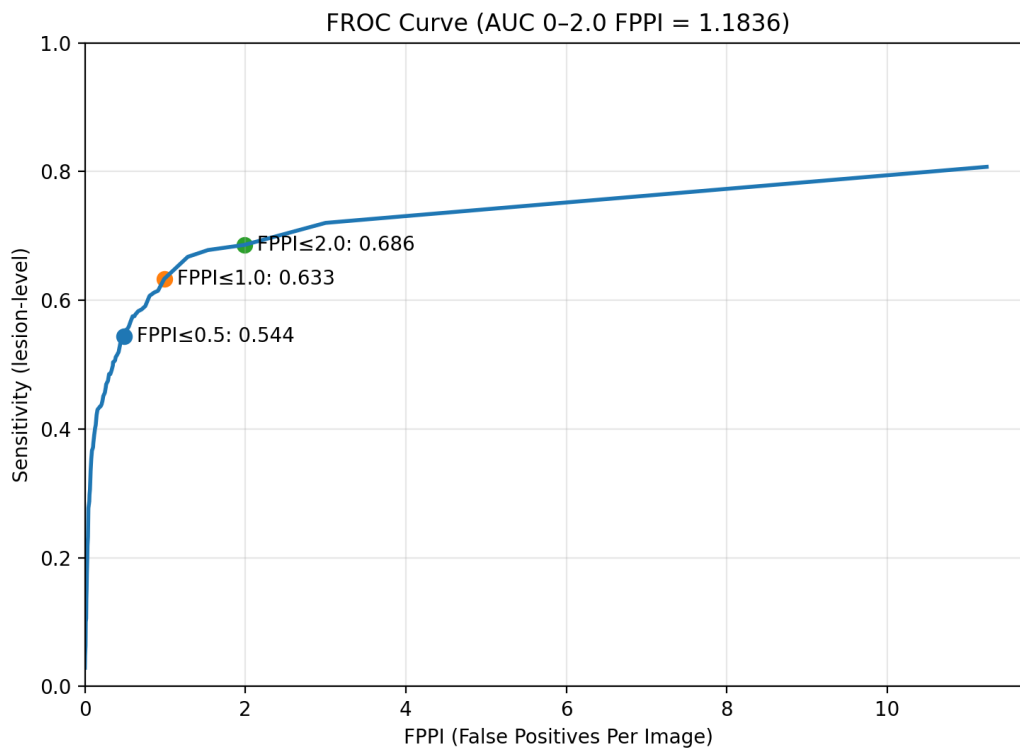


Figura 4.5: Curva FROC (sensibilidad vs falsos positivos por imagen, FPPI) en el conjunto de prueba.

Resultados destacados. Con base en la distribución del Índice de Jaccard, se reporta el porcentaje de casos con solapamiento alto (por ejemplo, Jaccard \geq 0,60), lo cual indica una localización adecuada en una proporción significativa del conjunto de prueba. De manera complementaria, el análisis FROC permite reportar la sensibilidad del modelo para distintos niveles tolerables de falsas alarmas por imagen, aportando una medida más alineada con escenarios de apoyo clínico.

Conclusiones

En este trabajo de grado se desarrolló y documentó un sistema reproducible para la detección automática de masas en mamografías digitalizadas utilizando un detector *one-stage* de la familia YOLO. A partir del subconjunto de masas de CBIS-DDSM, se implementó un pipeline completo que abarca la descarga y organización de datos, el preprocesamiento en formato DICOM, la derivación de *bounding boxes* desde ROI masks y la generación de anotaciones en formato YOLO, dejando el conjunto preparado para entrenamiento, validación y prueba.

Desempeño del modelo y lectura práctica de las métricas estándar. El entrenamiento base con YOLOv8m preentrenado permitió obtener un punto de partida sólido y estable para el problema de una sola clase (*mass*). En validación se alcanzaron métricas estándar de detección ($P = 0.586$, $R = 0.544$, $mAP@0.5 = 0.533$ y $mAP@0.5:0.95 = 0.243$), que, si bien no son excepcionalmente altas en términos de *benchmarks* genéricos, resultan coherentes con la complejidad inherente de la mamografía (bajo contraste, superposición de tejido y variabilidad morfológica).

Importancia del Índice de Jaccard como indicador de localización clínica. Con el objetivo de reportar una medida más interpretable desde el punto de vista clínico, se evaluó la concordancia espacial entre predicción y anotación mediante el Índice de Jaccard (IoU), incluyendo un procedimiento explícito para manejar la presencia de múltiples masas en una misma imagen mediante emparejamiento uno-a-uno. En el conjunto de prueba, el 39.61 % de las imágenes presentó un $\text{IoU} \geq 0,60$ (143 de 361), lo cual sugiere que, en una fracción relevante de casos, el sistema no solo *detecta* zonas de interés, sino que además *localiza* con una precisión espacial suficiente como para servir de apoyo visual al especialista (p. ej., resaltando regiones candidatas a inspección).

Importancia de FROC (sensibilidad vs. FPPI) para un escenario de apoyo clínico. En tareas de detección en imagen médica, no basta con reportar precisión o mAP: también es crucial cuantificar el compromiso entre (i) detectar la mayor cantidad de lesiones verdaderas (sensibilidad) y (ii) limitar la cantidad de falsas alarmas que el usuario debe revisar por imagen (FPPI). En este trabajo, la curva FROC permitió analizar ese compromiso de forma directa y alineada con el uso clínico. Bajo límites de $\text{FPPI} \leq 0,5$, $\leq 1,0$ y $\leq 2,0$, se obtuvieron sensibilidades de 0.5435, 0.6332 y 0.6860, respectivamente, evidenciando que al permitir más falsas alarmas por imagen es posible aumentar la sensibilidad del sistema. Este análisis aporta un criterio práctico para seleccionar el umbral de operación del modelo según el nivel de tolerancia a falsas alarmas en un flujo de lectura.

Rol del post-procesamiento (Extra NMS) para métricas confiables y salidas interpretables. Para que IoU y FROC reflejen el comportamiento real del detector (y no artefactos por duplicación), fue necesario implementar un pipeline de post-procesamiento con *Extra NMS* orientado a eliminar cajas altamente superpuestas dentro de una misma región. Este paso resultó clave para (i) evitar que múltiples predicciones redundantes inflaran falsos positivos y (ii) estabilizar el cálculo de solapamiento al comparar predicciones con *ground truth*. En consecuencia, el Extra NMS no solo mejora la interpretabilidad visual de las predicciones, sino que también permite que las métricas adicionales reportadas sean comparables y consistentes con el objetivo clínico.

5.1. Limitaciones

Una limitación importante del planteamiento experimental es que el conjunto de prueba no incluye mamografías negativas (todas las imágenes contienen al menos una masa). Esto implica que una imagen sin detecciones se interpreta como un fallo crítico (falso negativo), y restringe el análisis de especificidad en escenarios reales donde muchas mamografías pueden no presentar hallazgos.

5.2. Trabajo futuro

Como líneas de continuación, se identifican oportunidades claras:

- **Mayor capacidad computacional y entrenamientos más robustos:** realizar experimentos sistemáticos con mayor resolución, mayor *batch size* efectivo, y estrategias de entrenamiento más extensas (o con búsqueda de hiperparámetros) para verificar si se puede mejorar la localización (IoU) y la sensibilidad bajo límites de FPPI.
- **Evaluación más cercana al escenario real:** incorporar mamografías negativas y reportar análisis adicionales orientados a tamizaje, donde el control de falsas alarmas por imagen cobra especial relevancia.
- **Generalización y validación externa:** evaluar el modelo en datos de distinta procedencia (otros dispositivos/centros) para analizar cambio de dominio y robustez.
- **Escalamiento a herramienta de apoyo:** evolucionar desde un modelo experimental hacia un aplicativo prototipo (p. ej., interfaz que muestre predicciones, umbral configurable y reporte automático de IoU/FROC), enfocado en apoyar la lectura clínica como sistema de segunda mirada (*second reader*).

5.3. Cierre

En conjunto, los resultados muestran que, aunque las métricas tradicionales no reflejan un rendimiento sobresaliente en términos absolutos, el sistema logra detectar y localizar una proporción

significativa de zonas de interés. En particular, la combinación de IoU (Jaccard) y FROC (sensibilidad vs. FPPI) aporta una evaluación más alineada con el posible uso clínico, ya que cuantifica simultáneamente calidad de localización y carga de falsas alarmas, dos factores esenciales para considerar su utilidad como guía computacional en la revisión de mamografías.

Bibliografía

- [1] Intelligent Ultrasound, “ScanNav Anatomy Peripheral Nerve Block,” 5 2024.
- [2] World Health Organization, “Breast cancer,” Aug. 2025. Accessed: 2025-11-06.
- [3] American College of Radiology, *ACR BI-RADS[®] Atlas: Breast Imaging Reporting and Data System*. Reston, VA: American College of Radiology, 5th ed., 2013.
- [4] D. B. Kopans, *Breast Imaging*. Philadelphia, PA: Lippincott Williams & Wilkins, 3rd ed., 2007.
- [5] D. R. Dance, S. Christofides, A. D. A. Maidment, I. D. McLean, and K.-H. Ng, *Diagnostic Radiology Physics: A Handbook for Teachers and Students*. Vienna: International Atomic Energy Agency (IAEA), 2014. Capítulos sobre mamografía digital.
- [6] G. Jocher, A. Chaurasia, and J. Qiu, “Ultralytics yolo.” <https://docs.ultralytics.com/>, 2023. Accedido: 2025-11-10.
- [7] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, “A curated mammography data set for use in computer-aided detection and diagnosis research,” *Scientific Data*, vol. 4, p. 170177, 2017.
- [8] M. Heath, K. Bowyer, D. Kopans, R. Moore, and W. P. Kegelmeyer, “The digital database for screening mammography,” in *Proceedings of the Fifth International Workshop on Digital Mammography*, pp. 212–218, Medical Physics Publishing, 2000.
- [9] The Cancer Imaging Archive, “Curated breast imaging subset of ddsd (cbis-ddsm).” <https://www.cancerimagingarchive.net/collection/cbis-ddsm/>, 2017. Accedido: 2025-11-10.
- [10] J. L. Jesneck *et al.*, “Optimized approach to decision fusion of heterogeneous data for breast cancer diagnosis,” *Academic Radiology / Med. Imaging related (PMC)*, 2006. Describe digitizers y resoluciones: 42, 43.5 y 50 micras/píxel.
- [11] N. Alam *et al.*, “Classification of microcalcification clusters in digital mammograms using a stack generalization based classifier,” *Journal of Imaging*, 2019. Resume digitizers usados en DDSM: 42, 43.5 y 50 micras/píxel.
- [12] K. H. Cha *et al.*, “Evaluation of data augmentation via synthetic images for improved breast mass detection,” *Journal of Medical Imaging*, 2020. Menciona tamaños de píxel variables en DDSM: 42, 43.5 y 50 μm .
- [13] DICOM Standard Browser (Innolitics), “Imager pixel spacing attribute.” Online. Define PixelSpacing como distancia física entre centros de píxel (mm).

- [14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2016.
- [15] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “Yolov4: Optimal speed and accuracy of object detection,” *arXiv:2004.10934*, 2020.
- [16] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, “Generalized intersection over union: A metric and a loss for bounding box regression,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 658–666, 2019.
- [17] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, “Distance-iou loss: Faster and better learning for bounding box regression,” in *AAAI Conference on Artificial Intelligence*, pp. 12993–13000, 2020.
- [18] N. Bodla, B. Singh, R. Chellappa, and L. S. Davis, “Soft-nms – improving object detection with one line of code,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 5561–5569, 2017.
- [19] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *CVPR*, pp. 580–587, 2014.
- [20] R. Girshick, “Fast r-cnn,” in *ICCV*, pp. 1440–1448, 2015.
- [21] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” in *NeurIPS*, pp. 91–99, 2015.
- [22] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “Ssd: Single shot multibox detector,” in *ECCV*, pp. 21–37, 2016.
- [23] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *ICCV*, pp. 2980–2988, 2017.
- [24] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” in *CVPR*, pp. 2117–2125, 2017.
- [25] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, “Path aggregation network for instance segmentation,” in *CVPR*, pp. 8759–8768, 2018.
- [26] X. Li, W. Wang, L. Wu, S. Chen, X. Hu, J. Li, Z. Tang, and J. Yang, “Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection,” in *NeurIPS*, 2020.
- [27] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, “Yolox: Exceeding yolo series in 2021,” in *arXiv:2107.08430*, 2021. Introduce SimOTA para asignación dinámica de positivos.

- [28] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft coco: Common objects in context,” in *ECCV*, pp. 740–755, 2014.
- [29] D. C. Edwards, C. E. Metz, and R. M. Nishikawa, “The free-response receiver operating characteristic (froc) curve: A method for comparing the performance of diagnostic systems,” *Medical Physics*, vol. 29, no. 8, pp. 1601–1610, 2002.
- [30] D. P. Chakraborty, “Jafroc analysis of observer performance in the froc paradigm,” *Medical Physics*, vol. 38, no. S2, pp. S106–S115, 2011.
- [31] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *MICCAI*, pp. 234–241, 2015.
- [32] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask r-cnn,” in *ICCV*, pp. 2961–2969, 2017.
- [33] W. Lotter *et al.*, “A deep learning system to improve the accuracy and efficiency of reading breast cancer screening mammograms,” *Nature*, vol. 577, pp. 89–94, 2021.
- [34] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, “Deep learning to improve breast cancer detection on screening mammography,” *Scientific Reports*, vol. 9, p. 12495, 2019.
- [35] G. Ribli, A. Horváth, Z. Unger, P. Pollner, and I. Csabai, “Detecting and classifying lesions in mammograms with deep learning,” *Scientific Reports*, vol. 8, p. 4165, 2018.
- [36] Ultralytics, “Explore ultralytics yolov8.” Documentation. Accedido: 2026-01-10.
- [37] Ultralytics, “Model training with ultralytics yolo (train mode).” Documentation. Accedido: 2026-01-10.
- [38] Ultralytics, “Coco dataset (detection) – overview.” Documentation. Accedido: 2026-01-10.
- [39] Ultralytics, “Configuration (train settings and defaults).” Documentation. Accedido: 2026-01-10.