



Pontificia Universidad
JAVERIANA
Cali

Implementación de técnicas de aprendizaje profundo para la detección automática de tejido cancerígeno en imágenes de histopatología.

Luis Guillermo Cortés Mojica (Código 8992245))
Johan Sneider Sánchez Torres (Código 8992674)
Carolina Silva Varela (Código 8993329)

Proyecto Aplicado para optar al título de
Máster en Ciencia de Datos

Director:
Julián Gil González PhD

Facultad de Ingeniería y Ciencias
Maestría en Ciencia de Datos
Santiago de Cali, Julio 14 de 2025

Índice general

1. Introducción	5
2. Definición del problema	6
2.1. Planteamiento del problema	6
2.2. Formulación del problema	7
3. Objetivos del proyecto	8
3.1. Objetivo general	8
3.2. Objetivos específicos	8
4. Marco teórico y Antecedentes	9
4.1. Antecedentes	9
4.2. Marco teórico	11
5. Metodología	25
5.1. Enfoque general	25
5.2. Conjunto de datos	25
5.2.1. Distribución por clases	28
5.3. Preprocesamiento de imágenes	28
5.4. Arquitecturas y backbones	29
5.5. Entrenamiento	29
5.6. Evaluación del rendimiento	31
5.7. Comparación de configuraciones	32
6. Resultados	33
6.1. Objetivo 1	33
6.2. Objetivo 2 y Objetivo 3	36
6.2.1. Procesamiento con imágenes crudas.	36
6.2.2. Procesamiento con imágenes filtradas.	39
7. Discusión	44
8. Conclusiones	46
9. Trabajos futuros	48
Anexos	52

Índice de figuras

4.1.	Modelo conceptual de una CNN. Tomado de (1)	19
4.2.	Arquitectura U-Net. Tomado de (2)	19
4.3.	Arquitectura LinkNet. Tomado de (3)	20
4.4.	Arquitectura FPN. Tomado de (4)	21
4.5.	Backbone EfficientNet-B3. Elaboración propia.	22
4.6.	Backbone EfficientNet-B7. Elaboración propia.	22
4.7.	Backbone ResNet-50. Elaboración propia.	23
5.1.	Metodología para Segmentación Semántica en Imágenes de Histopatología - Diagrama de flujo construcción del modelo.	26
5.2.	Imagen histopatológica de cáncer de mama.	27
5.3.	Pipeline de preprocesamiento adaptativo para imágenes histopatológicas.	30
6.1.	Comparación visual de preprocesamiento	35
6.2.	LinkNet-EfficientNet-B7 sin filtro	38
6.3.	Matriz de confusión: LinkNet-EfficientNet-B7 sin filtro	38
6.4.	LinkNet-EfficientNet-B7 con filtro	41
6.5.	Matriz de confusión: LinkNet-EfficientNet-B7 con filtro	41
6.6.	Predicciones LinkNet con EfficientNet-B7 entrenada con imágenes sin filtrar	43
1.	Arquitectura LinkNet con imágenes sin filtro	52
2.	Arquitectura FPN con imágenes sin filtro	53
3.	Arquitectura UNet++ con imágenes sin filtro	53
4.	Arquitectura LinkNet con imágenes filtradas	54
5.	Arquitectura FPN con imágenes filtradas	55
6.	Arquitectura UNet++ con imágenes filtradas	56
7.	Arquitectura LinkNet con imágenes filtradas	57
8.	Arquitectura FPN con imágenes filtradas	57
9.	Arquitectura UNet++ con imágenes filtradas	58
10.	Arquitectura LinkNet con imágenes filtradas	59
11.	Arquitectura FPN con imágenes filtradas	60
12.	Arquitectura UNet++ con imágenes filtradas	61

Índice de cuadros

5.1. Distribución de píxeles por clase	28
6.1. Comparación de métricas para elección de filtro.	33
6.2. Métricas de calidad del preprocesamiento.	34
6.3. Dice por arquitectura y <i>backbones</i> en imágenes crudas.	36
6.4. Dice por clase por arquitectura y <i>backbones</i> en imágenes crudas.	37
6.5. Dice por arquitectura y <i>backbones</i> en imágenes filtradas.	39
6.6. Dice por clase por arquitectura y <i>backbones</i> en imágenes filtradas.	40

Capítulo 1

Introducción

El cáncer de mama es una de las principales causas de mortalidad en mujeres a nivel mundial, con aproximadamente 1.38 millones de nuevos casos y medio millón de defunciones anuales siguiendo las estadísticas de la Organización Mundial de la Salud (OMS). La creciente incidencia de esta enfermedad causa la necesidad de desarrollar métodos eficaces para su detección y tratamiento temprano. A pesar de que los avances de las técnicas de detección y tratamiento han reducido la tasa de mortalidad, la mejora continua en las metodologías es fundamental para detectar la enfermedad de una manera más efectiva.

La histopatología sigue siendo el estándar de oro para el diagnóstico de cáncer de mama; esta técnica consiste en la extracción y análisis de tejido mamario bajo el microscopio para identificar células cancerígenas. Sin embargo, este método al ser susceptible a errores humanos y a la diferencia de criterios por parte de los patólogos, surge la necesidad de análisis rápidos y precisos que permitan realizar una exploración más detallada usando técnicas automatizadas como los modelos de aprendizaje profundo basados en Inteligencia Artificial.

El aprendizaje profundo ha demostrado tener potencial en el campo del procesamiento de imágenes médicas. Las redes neuronales convolucionales (CNN) se vuelven adecuadas para la segmentación de imágenes debido a su capacidad para identificar patrones en datos visuales. Estas redes han sido adoptadas especialmente en segmentación de imágenes médicas, donde cada píxel de una imagen se clasifica en una categoría específica.

Este proyecto se centra en el desarrollo de un modelo de segmentación semántica para la detección automática de tejido cancerígeno en imágenes de histopatología de cáncer de mama. La implementación de este modelo se realiza con la intención de mejorar la precisión y eficacia del diagnóstico, facilitando así la detección temprana de la enfermedad y por ende mejorar el tiempo de respuesta al tratamiento, por tanto, se espera que el proyecto contribuya al conocimiento del uso de técnicas de aprendizaje profundo en el área clínica.

Capítulo 2

Definición del problema

2.1. Planteamiento del problema

El cáncer de mama es la neoplasia más prevalente entre las mujeres a nivel mundial. Esta patología se caracteriza por el crecimiento descontrolado de células malignas en el tejido mamario, formando tumores que, si no son detectados y tratados de forma oportuna, pueden causar metástasis y eventualmente la muerte. Según datos de la OMS, anualmente se registran aproximadamente 1.38 millones de casos nuevos y medio millón de muertes asociadas (5). A pesar de los avances en técnicas de diagnóstico y tratamiento, la creciente incidencia, que aumenta alrededor del 0.3 % anual, exige una mejora constante de los métodos diagnósticos, especialmente aquellos basados en tecnologías computacionales. Además, el diagnóstico tradicional por histopatología, aunque considerado el estándar de oro, implica una alta carga de trabajo, interpretación subjetiva y riesgo de error humano, lo que plantea una necesidad crítica de herramientas automatizadas y precisas que complementen el análisis clínico (6).

Desafortunadamente, el aumento de la tasa de incidencia de este tipo de cáncer es inevitable, pero con los avances actuales como el uso de la inteligencia artificial en el diagnóstico y elaboración de medicamentos, y la implementación de la criomicroscopía para la planificación de diferentes tratamientos, es posible disminuir el número de muertes causadas por la enfermedad. Sin embargo, para que esto sea posible, el diagnóstico y el tratamiento oportuno son cruciales, pues en etapas avanzadas la enfermedad tiende a hacer metástasis. Por lo tanto, la detección temprana de la enfermedad es de vital importancia para mejorar la salud y la expectativa de vida de los pacientes (6).

Actualmente, la histopatología sigue siendo el referente para el diagnóstico del cáncer a pesar de que existan diversas técnicas para su detección y diagnóstico. Este método consiste en extraer tejido infectado de la mama y examinarlo bajo el microscopio. Sin embargo, el porta objetos de histopatología contiene patrones visuales intrincados que son difíciles de categorizar como benignos o malignos, con lo cual se torna una tarea compleja y que demanda mucho tiempo a los especialistas que analizan las imágenes (7). Por esta razón, en la actualidad el uso de técnicas de aprendizaje profundo es cada vez más frecuente en diversas investigaciones médicas. No obstante, muchos de los algoritmos desarrollados se construyen y entrenan

a partir de los conjuntos de datos.

Así pues, la recolección, limpieza y preparación de los datos resulta ser una tarea fundamental que demanda una gran cantidad de recursos. Por otra parte, usualmente los algoritmos implementados exigen gran capacidad de cómputo y tienden a ser demasiado complejos al momento del despliegue.

2.2. Formulación del problema

La detección oportuna y precisa del cáncer de mama a partir de imágenes histopatológicas constituye un reto central en el ámbito clínico. La complejidad visual de estas imágenes, sumada a la necesidad de interpretación especializada y al riesgo de variabilidad interobservador, dificulta el diagnóstico rápido y fiable. Ante esta problemática, surge la necesidad de desarrollar sistemas automatizados basados en técnicas de aprendizaje profundo que permitan segmentar de forma precisa las regiones de tejido cancerígeno. Así, este trabajo se orienta a resolver la siguiente pregunta de investigación:

¿Cuál es el mejor modelo con técnicas de aprendizaje profundo en segmentación semántica para la detección automática de tejido cancerígeno en imágenes de histopatología?

Para lograrlo, se plantearon las siguientes preguntas:

- ¿Cuáles técnicas de preprocesamiento de imágenes histopatológicas de cáncer de mama permiten mejorar la calidad de las imágenes del conjunto de datos?
- ¿Qué técnicas de procesamiento de imágenes se pueden aplicar para desarrollar y entrenar modelos de segmentación semántica basados en técnicas de Deep Learning?
- ¿Cuál de los modelos implementados para la detección de células con cáncer de mama ofrece el mejor rendimiento teniendo en cuenta las métricas de desempeño?

Capítulo 3

Objetivos del proyecto

3.1. Objetivo general

Implementar un modelo de segmentación semántica usando técnicas de aprendizaje profundo para la detección automática de tejido cancerígeno en imágenes de histopatología.

3.2. Objetivos específicos

- Realizar preprocesamiento de imágenes histopatológicas de cáncer de mama.
- Entrenar modelos de segmentación semántica basados en técnicas de Deep Learning.
- Evaluar el rendimiento de los modelos de segmentación semántica desarrollados.

Capítulo 4

Marco teórico y Antecedentes

4.1. Antecedentes

A Review of Medical Image Segmentation Algorithms.

En el artículo de investigación “A Review of Medical Image Segmentation Algorithms” (8), publicado en Abril de 2021, se presenta una revisión y explicación de varios modelos de segmentación de imágenes tradicionales y también un modelo novedoso llamado “Lattice Boltzmann method (LBM)”, el cual aumenta la velocidad de cómputo durante la segmentación de imágenes, presentando exactitud y especificidad de más del 95 %. En este artículo se concluye que el modelo LBM presenta los beneficios de la adaptabilidad y la velocidad de proceso con alta calidad de imágenes y utilizando una cantidad razonable de recursos computacionales.

Semantic Segmentation of Pancreatic Cancer in Endoscopic Ultrasound Images Using Deep Learning Approach.

Con el fin de encontrar métodos más precisos y eficientes para la detección de cáncer de páncreas, el estudio “Semantic Segmentation of Pancreatic Cancer in Endoscopic Ultrasound Images Using Deep Learning Approach” (9) desarrolló metodología de Deep Learning para la segmentación semántica de este tipo de cáncer en imágenes de ultrasonido endoscópico (EUS). El estudio contempló 330 imágenes adquiridas de 150 pacientes que contaban con este diagnóstico durante enero de 2010 y julio de 2021 en el Hospital Universitario Inah en Corea del Sur. Se usaron técnicas de preprocesamiento de datos como el aumento de las imágenes a 256x256 píxeles para entrenar y probar el modelo de aprendizaje profundo, además de técnicas como rotación, escalado y deformación elástica para el mejoramiento en la calidad de las imágenes. Entre los resultados se encontró que el mejor método para segmentación de imágenes de cáncer de páncreas fue la Red de Características de Atención Profunda (DAF-Net), pues demostró un mejor rendimiento con un coeficiente de similitud DSC del 82.4 %, una Intersección Sobre Unión (IoU) del 72.3 %, un Área Bajo la Curva (AUC) de un 92.7 % para una sensibilidad del 89 % y especificidad del 98.1 % y finalmente una precisión del 85.1 %, lo que indica una buena eficacia para segmentar con precisión las regiones de cáncer de páncreas en imágenes EUS.

A Solution to the stagnation of multi-verse optimization: An efficient method for breast cancer pathologic images segmentation.

El algoritmo Optimizador Multiverso, MVO por sus siglas en inglés, ha sido utilizado

en segmentación de imágenes, selección de características, problemas de ingeniería y en muchos otros campos. En el artículo “A Solution to the stagnation of multi-verse optimization: An efficient method for breast cancer pathologic images segmentation” (10). Publicado en 2023, se presenta una extensión al algoritmo MVO llamada CBQMVO mediante tres estrategias: 1) Adaptación de la estrategia de la matriz de covarianza (CMAES), 2) Estrategia de aprendizaje basada en biogeografía (BLS) y 3) Estrategia de cuasi-oposición y cuasi-reflexión (QROS). En este experimento se realizan comparaciones con diferentes métodos reconocidos y se concluye que el método CBQMVO es bastante competitivo y mejora al método MVO. Adicionalmente invitan a continuar con investigaciones sobre este modelo, pues, el CBQMVO puede aún resolver muchos problemas de ingeniería, señales médicas, energías renovables, entre otros.

Breast Cancer Histopathological Images Segmentation Using Deep Learning.

En el artículo “Breast Cancer Histopathological Images Segmentation Using Deep Learning” (11) se proponen dos enfoques para la segmentación semántica de imágenes histopatológicas de cáncer de mama. Por un lado, se propone una arquitectura de autocodificador para la segmentación no supervisada, y por otro lado, se propone una arquitectura U-Net de mejora para la segmentación supervisada. El enfoque no supervisado propuesto segmenta los núcleos en dos grupos basados en un modelo de autocodificadores, que les permite aprovechar los datos no etiquetados. Por su parte, para el enfoque supervisado se usó una arquitectura U-Net mejorada, que es un enfoque supervisado que requiere datos de entrenamiento etiquetados, pero ofrece una arquitectura más especializada para las tareas de segmentación. Los resultados experimentales evaluados en conjuntos de imágenes histopatológicas con diferentes métodos de tinción de color muestran que el método no supervisado puede segmentar eficazmente la imagen histológica. En particular, proporciona un rendimiento de segmentación comparativo con un modelo U-Net en términos de segmentación de núcleos. De acuerdo con las métricas de desempeño, el modelo U-Net propuesto y el modelo de segmentación no supervisada mostraron un rendimiento del 86.1% y 84.8%, respectivamente. Finalmente, los resultados obtenidos con la investigación muestran que estos modelos son más robustos en comparación con los métodos de la literatura.

Crowdsourcing Segmentation of Histopathological Images Using Annotations Provided by Medical Students

Teniendo en cuenta que los modelos de aprendizaje automático requieren de una gran cantidad de datos etiquetados. Una posible solución es la implementación de una técnica conocida como crowdsourcing, en la cual el etiquetado de las imágenes es distribuido en un grupo de individuos no necesariamente expertos. En el artículo “Crowdsourcing Segmentation of Histopathological Images Using Annotations Provided by Medical Students” (7) se presenta un nuevo enfoque del crowdsourcing para la segmentación de imágenes de cáncer de mama triple negativo, que consiste en la implementación de una red neuronal convolucional basada en una arquitectura U-Net que incorpora elementos de la arquitectura ResNet-34 debido al comportamiento ruidoso de los anotadores. El modelo fue aplicado a un conjunto de datos del mundo real, donde se pudo observar que este método logra un nivel de precisión

del 0.7578, el cual supera a otros modelos de segmentación y se acerca al resultado obtenido usando datos etiquetados por los expertos (0.7723).

4.2. Marco teórico

Cáncer de mama:

Según (12), cuando las células sanguíneas de la mama se vuelven recalcitrantes, se habla de cáncer de mama. Existen varios tipos de cáncer de mama, según las células afectadas; las tres partes principales de las mamas son los lobulillos, los conductos y el tejido conjuntivo. Los lobulillos son donde se produce la leche, mientras que los conductos son canales que llevan la leche al pezón. El tejido conjuntivo rodea y mantiene todo unido. La mayoría de los cánceres se producen en los lobulillos o conductos y pueden extenderse a otros órganos del cuerpo a través de los vasos sanguíneos. Los tipos más comunes de cáncer de mama son los siguientes:

- **Carcinoma ductal invasivo:** En este tipo de cáncer, las células cancerosas se producen fuera de los conductos.
- **Carcinoma lobulillar invasivo:** En este tipo de cáncer, las células cancerosas se producen fuera de los lobulillos.

Existen otros tipos de cáncer de mama como el medular de icono externo, el mucinoso y el inflamatorio. Los síntomas del cáncer de mama varían de una paciente a otra. Algunos pacientes pueden no tener todos los síntomas o ningún síntoma (12).

Cáncer de mama triple negativo:

Los tipos más comunes de receptores como el estrógeno, la progesterona y el gen HER2/neu, conocidos por impulsar el crecimiento del cáncer de mama, no se presentan en los tumores. Esto significa que el análisis de las células cancerosas es negativo para varias pruebas de laboratorio hormonales para HER2, estrógeno y progesterona. Para el tratamiento de este tipo de cáncer de mama se utiliza mucho la quimioterapia. Este tipo de cáncer se da en torno al 10-20% de los casos de cánceres de mama diagnosticados. Este tipo de cáncer se da sobre todo en mujeres jóvenes, afroamericanas, hispanas y/o con una mutación en el gen BRCA1. Este tipo de cáncer es más agresivo y difícil de tratar (12).

Segmentación de imágenes:

La segmentación de imágenes consiste en dividir una imagen en diferentes regiones o categorías que corresponden a diferentes objetos o partes de ellos, en donde cada región contiene píxeles con características similares (13). Se realiza la segmentación con la finalidad de identificar objetos y hacer más fácil su análisis y procesamiento.

La segmentación de imágenes es utilizada en múltiples áreas como visión por computadora, en automóviles para conducción autónoma, también es ampliamente usada en medicina para análisis y procesamiento de imágenes que permiten ayudar en el diagnóstico en general de enfermedades.

Métodos de segmentación de imágenes:

Para la segmentación de imágenes se pueden utilizar diferentes técnicas, dentro de las cuales se encuentran las siguientes (13):

- Umbralización: Consiste en definir un pixel como umbral para crear una imagen binaria en escala de grises generalmente usada para poder diferenciar si cada pixel hace parte del objeto o del fondo de la imagen. Se puede utilizar un valor único de umbral o definir varios umbrales para detectar clases. La forma básica de umbralización se define matemáticamente como:

$$g(x, y) = \begin{cases} 1 & \text{si } f(x, y) \geq T \\ 0 & \text{si } f(x, y) < T \end{cases}$$

Donde: $f(x, y)$ es la imagen de entrada en escala de grises, $g(x, y)$ es la imagen binaria resultante y T es el valor de umbral

- Segmentación basada en regiones: Esta segmentación se basa en un pixel y los vecinos que lo rodean, dentro de esta segmentación se pueden encontrar diferentes tipos, como son crecimiento de regiones, división, fusión y Watershed. Una segmentación basada en regiones se define como una partición de la imagen I en n regiones disjuntas:

$$I = \bigcup_{i=1}^n R_i$$

Donde se cumplen las siguientes propiedades:

- $R_i \cap R_j = \emptyset, \quad \forall i \neq j$
 - $\bigcup_{i=1}^n R_i = I$
 - $P(R_i) = \text{True}, \quad \forall i = 1, 2, \dots, n$
 - $P(R_i \cup R_j) = \text{False}, \quad \forall i \neq j$ con R_i y R_j adyacentes
- Segmentación por agrupación: Se juntan datos homogéneos en clusters utilizando el algoritmo k-medias, en este modelo, los componentes del set de datos pertenecen solo a una agrupación a la vez. La segmentación por clustering se formula como un problema de partición óptima del conjunto de píxeles:

$$\mathcal{P}^* = \arg \min_{\mathcal{P}} \sum_{i=1}^k \sum_{\mathbf{x} \in C_i} d(\mathbf{x}, \boldsymbol{\mu}_i)$$

Donde: $\mathcal{P} = C_1, C_2, \dots, C_k$ es una partición del espacio de características, $\boldsymbol{\mu}_i$ es el centroide del cluster C_i , $d(\cdot, \cdot)$ es una función de distancia y k es el número de clusters

- Segmentación binaria: es una técnica de procesamiento de imágenes que divide una imagen en exactamente dos categorías o clases:
 - Blanco (o valor 1) = píxeles que pertenecen al objeto que queremos identificar

- Negro (o valor 0) = píxeles que pertenecen al fondo
- Segmentación semántica: Es un algoritmo de aprendizaje profundo en el cual se asigna una etiqueta a cada píxel para definir su clase. Esta es muy útil en procesamiento de imágenes debido a que genera un nivel de detalle muy alto, sin embargo, los requerimientos computacionales son altos. La segmentación semántica se formula como un problema de mapeo píxel a píxel:

$$f : R^{H \times W \times C} \rightarrow \{1, 2, \dots, K\}^{H \times W}$$

Donde: $H \times W \times C$ representa una imagen de entrada con altura H , anchura W y C canales; K es el número de clases semánticas; la función f asigna cada píxel (i, j) a una clase $k \in 1, 2, \dots, K$

Preprocesamiento de imágenes:

El preprocesamiento de imágenes es un conjunto de técnicas aplicadas para mejorar la calidad de ellas previo a su análisis. El objetivo es transformar los datos de imágenes crudas en datos más limpios y útiles. Las imágenes originales suelen contener ruido y valores faltantes, inconsistentes o erróneos lo que puede afectar su análisis. La corrección de estos problemas es importante para mejorar la calidad de la detección de características y análisis de la imagen (14).

1. **Corrección de imágenes:** Las correcciones de imagen se agrupan generalmente en correcciones radiométricas y geométricas. Las primeras se enfocan en corregir errores causados por la irregularidad en la respuesta del sensor a lo largo de la imagen, mientras que las segundas abordan las distorsiones causadas por la geometría del sensor, la plataforma de captura o el movimiento de la cámara. La corrección radiométrica incluye la eliminación de errores como el ruido de píxeles aleatorios, la caída de líneas o columnas y los problemas de sincronización en la captura de imágenes. La corrección geométrica, por otro lado, implica la remuestreo y transformación de las coordenadas de la imagen para ajustarlas a una representación correcta en el mundo real.
2. **Mejora de imágenes:** La mejora de imágenes tiene como objetivo refinar la sensibilidad de la información en las imágenes, mejorando su entrada para otros métodos automáticos de procesamiento de imágenes. Este proceso puede involucrar la modificación de los atributos de la imagen, como el contraste o la nitidez, para hacer que los detalles importantes sean más visibles. Existen métodos de mejora en el dominio espacial, que alteran directamente los valores de los píxeles, y en el dominio de frecuencia, que trabajan con las transformaciones de la imagen.

Dentro de los métodos para mejora de las imágenes se encuentran:

- Transformación Power Law: También conocida como corrección gamma, se utiliza para mejorar la calidad de la imagen. Este tipo de corrección es utilizado por el cerebro humano, pantallas de televisión y monitores para

mostrar el mejor contraste de las imágenes. La transformación Power Law se define matemáticamente como:

$$s = c \cdot r^\gamma$$

Donde: s es la intensidad de salida (transformada) r es la intensidad de entrada (original), c es una constante de escalamiento positiva y γ (gamma) es el exponente que controla la forma de la transformación.

- Ecuación de histograma adaptativo limitado por contraste (CLAHE): La imagen se divide en pequeñas regiones y el histograma de cada región es computado. Luego se elige un límite de contraste como umbral. Se elige un límite de contraste como umbral para recortar el histograma en cada contenedor, y los píxeles por encima del umbral no se ignoran, sino que se distribuyen a otros contenedores antes de aplicar la ecualización del histograma en cada subimagen, finalmente se aplica una interpolación bilineal para remover artefactos en los límites de las subimágenes. En este caso el CLAHE se define como:

$$I_{\text{enhanced}}(x, y) = T_{(i,j)}[I_{\text{original}}(x, y)]$$

Donde $T_{(i,j)}$ es la función de transformación específica para el tile (i, j) que contiene el píxel (x, y) .

3. **Eliminación de ruido:** La eliminación de ruido de imágenes es un proceso en el que se emplean varios filtros para eliminar el ruido de cualquier imagen. Una técnica estándar de eliminación de ruido debe ser capaz de producir una imagen de alta calidad sin alterar los detalles de la imagen original. Estos detalles minuciosos son cruciales, especialmente en imágenes de tomografía computarizada o cualquier otra imagen médica. Los filtros son las técnicas más comunes empleadas para eliminar el ruido. Las herramientas que se utilizan para eliminar el ruido de cualquier imagen médica se denominan filtros. Los filtros se clasifican en lineales y no lineales (15), como se describen a continuación.

- Filtrado espacial de imágenes: Según (16), una de las principales técnicas que componen el procesamiento digital de imágenes es la de realce o mejora de imágenes. Esta técnica comprende un conjunto de operaciones tendientes a mejorar la calidad visual de una imagen. Estas operaciones permiten realzar las características de brillo y contraste de una imagen, reducir su contenido de ruido, o agudizar o intensificar detalles presentes en ella. En esta técnica, tal como en otras operaciones de procesamiento digital de imágenes, intervienen una imagen de entrada (input image) y una imagen de salida (output image). La primera constituye la imagen cuyos datos serán procesados, es decir, sometidos al realce, y la segunda es la resultante de tal procesamiento. Las operaciones que componen la técnica de realce pueden dividirse en dos tipos:

1. Operaciones de procesamiento puntual o pixel por pixel: estas operaciones tienden a mejorar el contraste tonal en la imagen, esto es, la diferencia entre los valores más oscuros y más claros que se visualizan

en un monitor. El procesamiento puntual altera los niveles de gris de los píxeles de una imagen. El nivel de gris de cada píxel en la imagen de entrada es modificado por un nuevo valor mediante operaciones matemáticas o relaciones lógicas. El valor resultante es ubicado en la imagen de salida en la misma posición espacial, esto es, en la misma posición (x, y) del arreglo rectangular de píxeles correspondiente a la imagen de entrada, donde x indica columna e y fila. Esto se realiza pixel por pixel, de manera individual, y los píxeles vecinos no tienen influencia (16).

2. Operaciones de procesamiento por grupo de píxeles o sobre vecindades: estas operaciones mejoran el contraste espacial en la imagen, es decir, la diferencia entre el valor digital de brillo de un determinado píxel y el de sus vecinos. Pretenden suavizar o reforzar estos contrastes espaciales de forma tal que, los valores de brillo en cada pixel de la imagen se asemejen o diferencien más de los correspondientes a los píxeles que los rodean. El procesamiento por grupo de píxeles opera sobre un grupo de píxeles de entrada que circundan a un píxel central. Los píxeles vecinos proveen información valiosa sobre las tendencias del brillo en el área de procesamiento (16). Para una imagen $I(x, y)$, el filtro de mediana se define como:

$$I_{\text{filtered}}(x, y) = \text{median}\{I(i, j) : (i, j) \in W(x, y)\}$$

- Filtrado de la mediana: El filtrado de mediana es un método de filtrado digital no lineal que se utiliza para eliminar el ruido de las imágenes. Se utiliza mucho en el procesamiento digital de imágenes porque preserva los bordes a la vez que elimina el ruido. Este filtro es eficaz para eliminar el ruido de tipo «sal y pimienta». Este filtro funciona recorriendo la imagen píxel a píxel y sustituyendo cada valor por el valor medio de los píxeles vecinos. Este filtrado es un tipo de técnica de suavizado, similar al filtrado lineal gaussiano. La mayoría de las técnicas de suavizado son eficaces para eliminar el ruido en las partes suaves de una señal, pero afectan principalmente a los bordes. Por eso, el filtrado de mediana se utiliza mucho en el tratamiento digital de imágenes. Por ejemplo, podemos filtrar una imagen para resaltar determinadas características o eliminar otras. Las operaciones de tratamiento de imágenes que se realizan con el filtrado son principalmente el suavizado, la nitidez y la mejora de los bordes. Las dos principales ventajas del filtrado de la mediana sobre el filtro de la media es que la mediana es muy robusta en comparación con la media y, por lo tanto, un único píxel muy poco representativo en un vecindario no afectará significativamente al valor de la mediana (16).
- Filtrado Wiener: El filtro de Wiener es la técnica más importante utilizada para eliminar la distorsión debida al movimiento lineal o a la óptica desenfocada en las imágenes. El objetivo principal de este filtro es eliminar el ruido que ha corrompido una señal. Es de naturaleza estadística. Reduce el MSE (error cuadrático medio) en la medida de lo posible. Los filtros de Wiener se utilizan para eliminar el ruido poisson y el ruido speckle. También se denomina filtrado inverso y es muy sensible al ruido aditivo. La estimación lineal de la imagen original se realiza mediante

este filtro (16).

- **Filtrado Gaussiano:** La función gaussiana o una aproximación a ella es una respuesta del filtro gaussiano relacionada con la electrónica y el tratamiento de señales. En el tratamiento de imágenes, es un filtro de paso bajo lineal y no uniforme. Se utiliza para desenfocar la imagen o reducir el ruido. El diseño del filtro gaussiano puede controlarse manipulando una sola variable, la varianza (16).

El filtrado gaussiano se utiliza en muchas áreas de investigación, como en la definición de la distribución de probabilidad del ruido o los datos, en el operador de suavizado “and” en matemáticas. En este filtrado, el píxel ruidoso puede sustituirse por el valor resultante de la multiplicación de la matriz del núcleo y la región seleccionada de la imagen. El filtro gaussiano es más eficaz en el suavizado de imágenes y se utiliza básicamente en el sistema de percepción visual humano. Es el primer paso en la detección de bordes, pero no es especialmente eficaz para eliminar el ruido de sal y pimienta (16).

- **Filtrado de la media:** El filtro medio o promedio es un filtro espacial de ventana deslizante lineal simple en el que el valor del píxel central se sustituye por la media de los valores de los píxeles de sus vecinos inmediatos (incluido el píxel central). Así se suavizan los impulsos o las grandes variaciones de valor de sus vecinos. Podemos tomar la media aritmética o geométrica de sus vecinos (16).

La principal ventaja del filtro de medias es que es muy sencillo conceptualmente y su aplicación es fácil. Puede ser muy útil para reducir el ruido impulsivo (o ruido de sal y pimienta) y el ruido gaussiano. Sin embargo, provoca un poco de desenfoque, lo que se traduce en la pérdida de detalle tras el filtrado (16).

- **Filtrado de Non-Local Means (NLM):** es un algoritmo de reducción de ruido en imágenes digitales introducido por Buades, Coll y Morel (17), que fundamenta su funcionamiento en el principio de promediar píxeles con características similares a nivel global en la imagen, en contraste con los métodos tradicionales que operan únicamente en vecindades locales, el filtro NLM explora similitudes estructurales en toda la imagen, aprovechando patrones repetitivos y texturas distribuidas globalmente.

El algoritmo evalúa la similitud entre píxeles comparando parches completos de imagen en lugar de valores individuales. Los pesos se asignan mediante un kernel exponencial que favorece píxeles con parches similares, permitiendo un promedio ponderado más inteligente que preserva estructuras importantes mientras elimina el ruido.

4. **El detector de bordes de Canny:** constituye uno de los algoritmos fundamentales en el procesamiento de imágenes digitales y visión por computadora. Desarrollado por John F. Canny en 1986 y presentado en su trabajo seminal (18), este método ha revolucionado el campo de la detección de bordes al introducir un enfoque matemáticamente riguroso y óptimo para la identificación

de contornos en imágenes.

La importancia del algoritmo de Canny radica en que se ha establecido como uno de los métodos más rigurosamente definidos que proporciona detección buena y confiable, siendo ampliamente utilizado en sistemas de visión por computadora. Su popularidad se debe a que cumple óptimamente con tres criterios fundamentales: buena detección, buena localización y respuesta única a un borde (19), características que lo convierten en una herramienta indispensable en aplicaciones de ciencia de datos que requieren análisis preciso de estructuras visuales.

- Suavizado con filtro Gaussiano: Para reducir el ruido en la imagen original $I(x, y)$, se aplica un filtro Gaussiano:

$$G_{\sigma}(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

La imagen suavizada $I_s(x, y)$ se obtiene por convolución:

$$I_s(x, y) = G_{\sigma}(x, y) * I(x, y)$$

- Cálculo del gradiente: Se calculan las derivadas parciales en las direcciones x e y :

$$G_x = \frac{\partial I_s}{\partial x}, \quad G_y = \frac{\partial I_s}{\partial y}$$

- Magnitud y dirección del gradiente: La magnitud del gradiente, que indica la fuerza del borde, es:

$$G = \sqrt{G_x^2 + G_y^2}$$

La dirección del gradiente es:

$$\theta = \arctan\left(\frac{G_y}{G_x}\right)$$

- Supresión no máxima y umbralización: Estos pasos se realizan mediante algoritmos para afinar los bordes y decidir qué píxeles corresponden a bordes reales basándose en dos umbrales T_{low} y T_{high} .

5. **Aumentación de datos:** La aumentación de datos busca introducir variabilidad controlada en el conjunto de entrenamiento para que el modelo aprenda invariancias frente a transformaciones que no afectan la semántica de la tarea. Esto conduce a:

- Mejor generalización, al exponer la red a múltiples versiones de cada imagen.
- Reducción del sobreajuste, al impedir que el modelo memorice detalles espurios de las muestras originales.

- Mayor robustez, ya que el modelo aprende a segmentar correctamente pese a ligeras distorsiones geométricas o de intensidad.

En el contexto de histopatología, los patrones de tejido suelen ser estructuralmente similares aunque rotados o volteados, lo cual hace que algunas transformaciones (por ejemplo el volteo horizontal) conserven la validez de la etiqueta.

Aprendizaje profundo:

El aprendizaje automático es hoy uno de los campos tecnológicos más importantes y de más rápido crecimiento. Las aplicaciones del aprendizaje automático se están volviendo omnipresentes, por lo que las soluciones aprendidas a partir de datos están desplazando cada vez más a los algoritmos tradicionales elaborados a mano. Esto no sólo ha llevado a un mejor rendimiento de las tecnologías existentes, sino que también ha abierto la puerta a una amplia gama de nuevas capacidades que serían inconcebibles si los nuevos algoritmos tuvieran que diseñarse explícitamente a mano (20).

Actualmente, una rama concreta del aprendizaje automático, conocida como aprendizaje profundo, ha surgido como un marco excepcionalmente potente y de uso general para aprender de los datos. El aprendizaje profundo se basa en modelos computacionales denominados redes neuronales, que se inspiraron originalmente en los mecanismos de aprendizaje y procesamiento de la información del cerebro humano. En este sentido, se puede entender al aprendizaje profundo como un subcampo de la inteligencia artificial que se centra en la creación de grandes modelos de redes neuronales que sean capaces de tomar decisiones precisas basadas en datos, siendo particularmente adecuado para contextos donde se dispone de grandes y complejos conjuntos de datos, estos algoritmos permiten tomar decisiones basadas en los datos, identificando y extrayendo patrones de grandes conjuntos de datos que mapean con precisión conjuntos de datos complejos para obtener buenos resultados en la toma de decisiones (21).

Redes Neuronales Convolucionales:

Las redes neuronales convolucionales (CNN) también llamadas como ConvNets, son un tipo de red neuronal artificial (ANN) con una arquitectura de retroalimentación profunda. Son conocidas por su capacidad de generalización, superando a otras redes con capas completamente conectadas (FC) en la identificación y clasificación de características abstractas, especialmente en datos espaciales como imágenes. Un modelo profundo de CNN se compone de múltiples capas de procesamiento, cada una es capaz de extraer diferentes características de los datos de entrada a distintos niveles de abstracción. Las primeras capas se centran en detectar características de nivel superior con menor abstracción, mientras que las capas posteriores extraen características de nivel inferior con mayor abstracción (22). En la Figura 4.1 se expone el modelo conceptual de una CNN.

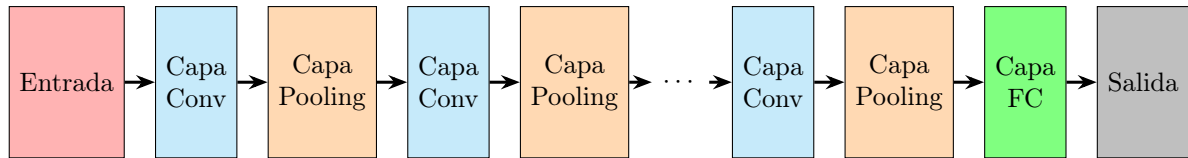


Figura 4.1: Modelo conceptual de una CNN. Tomado de (1)

Arquitectura U-Net++:

La UNet++ constituye una evolución significativa de la arquitectura U-Net tradicional, está específicamente diseñada para abordar las limitaciones inherentes en la segmentación de imágenes médicas.

La arquitectura UNet++ se fundamenta en el principio de que las redes convolucionales se pueden optimizar de manera más eficiente cuando los mapas de características del codificador y decodificador mantienen similitud semántica. Esta premisa contrasta con las conexiones de salto convencionales de la U-Net, que fusionan directamente características semánticamente disímiles entre las sub-redes codificadoras y decodificadoras (23).

Finalmente, la arquitectura U-Net viene definida como se muestra en la Figura 4.2.

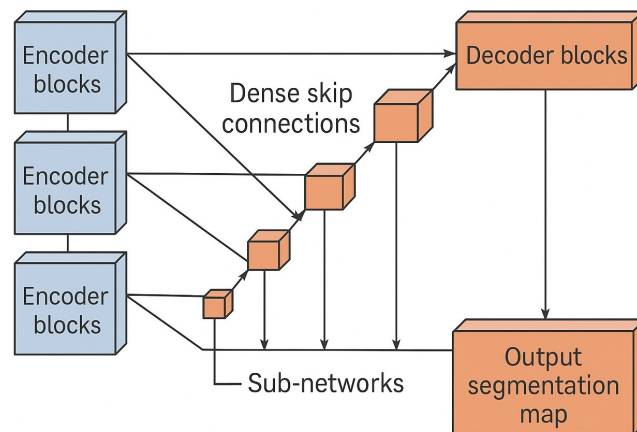


Figura 4.2: Arquitectura U-Net. Tomado de (2)

Arquitectura LinkNet:

Es una arquitectura diseñada específicamente para segmentación semántica eficiente en tiempo real (3). La arquitectura se compone de un codificador (encoder) que reduce la resolución espacial de la imagen y un decodificador (decoder) que la reconstruye, aplicando segmentación a nivel de píxel, como se observa en la Figura 4.3. Lo distintivo de LinkNet es el uso de conexiones directas (skip connections) entre cada bloque del encoder y su bloque correspondiente en el decoder, permitiendo una recuperación más precisa de detalles espaciales sin un gran costo computacional.

Estructuralmente, el encoder utiliza bloques convolucionales inspirados en arquitecturas como ResNet, mientras que el decoder realiza operaciones de upsampling para restaurar la resolución espacial original. A diferencia de UNet, que concatena las

salidas del encoder y decoder, LinkNet realiza una suma directa, lo que reduce el número de parámetros sin afectar significativamente la precisión.

Entre sus ventajas se destacan su bajo costo computacional, su flexibilidad para utilizar distintos backbones, y su adecuación a tareas en tiempo real. LinkNet ha sido usada exitosamente en contextos como imágenes médicas, conducción autónoma, agricultura de precisión y análisis de imágenes satelitales.

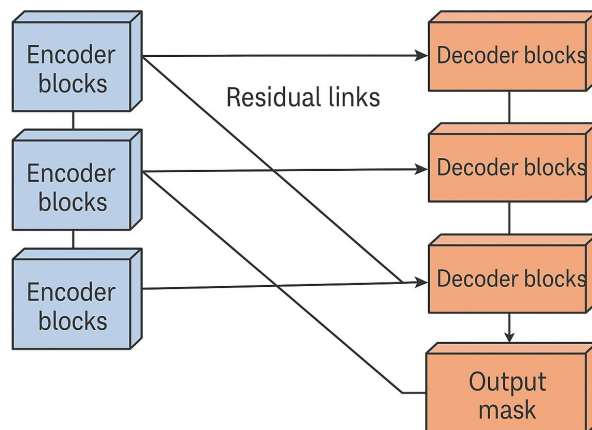


Figura 4.3: Arquitectura LinkNet. Tomado de (3)

Arquitectura Feature Pyramid Network (FPN):

La Arquitectura de Pirámide de Características (FPN) constituye un avance significativo en el ámbito de la visión por computadora, particularmente para abordar el desafío de la detección multi-escala de objetos. Desarrollada originalmente por Lin et al. (4), esta estructura combina de manera eficiente dos flujos de procesamiento: un camino ascendente (bottom-up), encargado de extraer características jerárquicas mediante redes convolucionales profundas (como ResNet o VGG), y un camino descendente (top-down), que reconstruye progresivamente la resolución espacial mediante operaciones de upsampling. La sinergia entre ambos caminos se logra mediante conexiones laterales, las cuales integran mapas de características de diferentes niveles mediante convoluciones 1×1 y fusiones aditivas.

Esta integración permite preservar información semántica de alto nivel en capas profundas, mientras se mantienen detalles espaciales finos en capas superficiales, superando así las limitaciones de las pirámides de imagen tradicionales. A diferencia de métodos anteriores, que requerían procesar múltiples escalas de entrada por separado, la FPN genera una pirámide de características en una sola pasada, reduciendo costos computacionales y facilitando su implementación en sistemas en tiempo real. Además, su diseño modular permite acoplar cabezales de predicción compartidos para tareas como detección o segmentación, lo que mejora la consistencia en resultados a diferentes escalas.

Las aplicaciones de la FPN abarcan desde sistemas de detección de objetos (por ejemplo, Faster R-CNN y RetinaNet) hasta análisis de imágenes satelitales para

identificación de cambios morfológicos. Su eficacia radica en la capacidad de generalizar contextos multi-escala sin sacrificar precisión, incluso en entornos con alta variabilidad de tamaño y posición de objetos, como en imágenes médicas o de vigilancia aérea. Su arquitectura se muestra en la Figura 4.4.

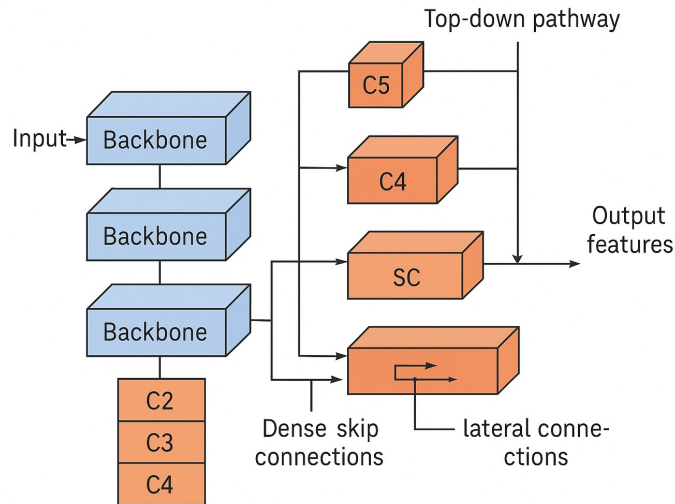


Figura 4.4: Arquitectura FPN. Tomado de (4)

Backbones:

En arquitecturas de segmentación semántica como U-Net, FPN, LinkNet o DeepLabV3+, el backbone o codificador es el componente encargado de extraer representaciones jerárquicas de la imagen. La calidad de estas representaciones influye directamente en la precisión de la segmentación. Entre los backbones más utilizados y eficientes se encuentran EfficientNet-B3, EfficientNet-B7 (disponibles a través de la biblioteca timm), y ResNet-50. A continuación, se describen sus fundamentos, ventajas y diferencias.

1. EfficientNet (B3 y B7): es una familia de redes convolucionales, que introducen un enfoque de scaling compuesto para optimizar simultáneamente la profundidad, el ancho y la resolución de entrada de la red. A diferencia de otras arquitecturas que aumentan arbitrariamente alguno de estos factores, EfficientNet escala los tres de manera balanceada, logrando mejor precisión y eficiencia computacional (24). La Figura 4.5 y la Figura 4.6 muestran las arquitecturas EfficientNet-B3 y EfficientNet-B7 respectivamente.
2. ResNet-50: forma parte de la familia de Redes Residuales introducida por He et al. (2015), diseñada para entrenar redes muy profundas sin que el rendimiento se degrade por el problema del desvanecimiento del gradiente. Esto se logra mediante conexiones residuales (skip connections) que permiten un flujo directo del gradiente durante el aprendizaje (23). La Figura 4.7 muestra la respectiva arquitectura.

Mecanismo de unsampling único: el decodificador utiliza índices de pooling calculados en el paso de max-pooling del codificador correspondiente para realizar un

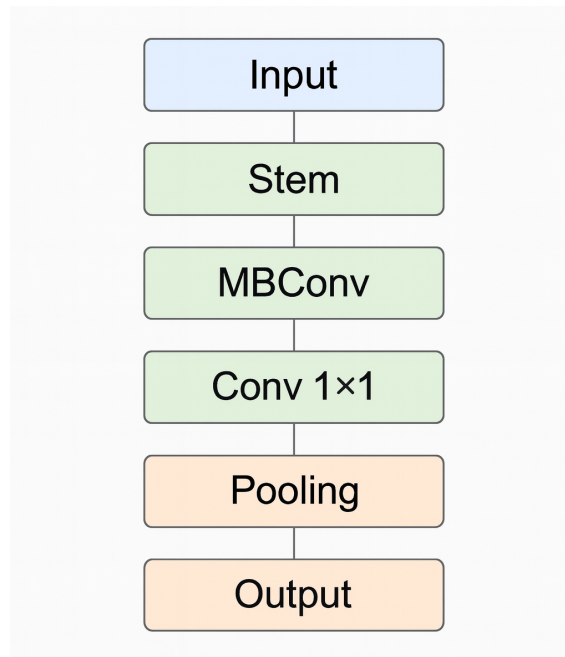


Figura 4.5: Backbone EfficientNet-B3. Elaboración propia.

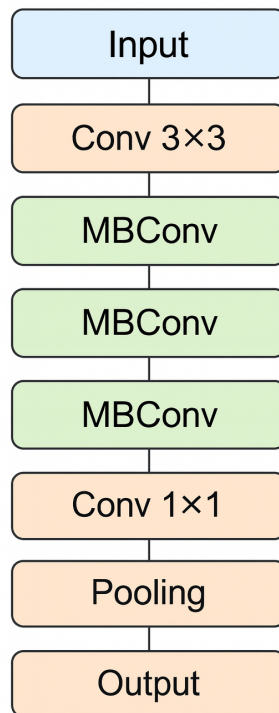


Figura 4.6: Backbone EfficientNet-B7. Elaboración propia.

upsampling no lineal Eficiencia computacional: Diseñada para ser eficiente tanto en términos de memoria como de tiempo computacional durante la inferencia, con significativamente menos parámetros entrenables que otras arquitecturas competidoras.

Peak Signal-to-Noise Ratio (PSNR)

Es una métrica objetiva ampliamente utilizada para evaluar la calidad de imágenes

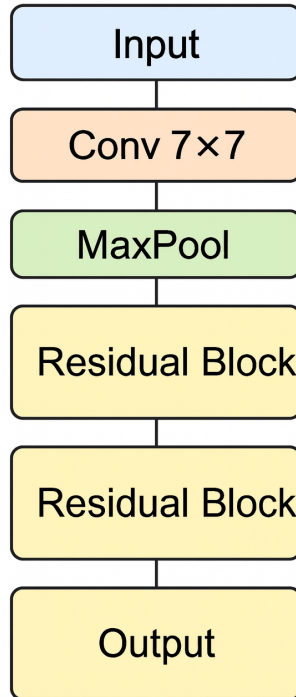


Figura 4.7: Backbone ResNet-50. Elaboración propia.

digitales procesadas o reconstruidas. Se define como la relación entre la máxima potencia posible de una señal y la potencia del ruido de corrupción que afecta la fidelidad de su representación (25). Esta métrica cuantifica numéricamente las diferencias entre una imagen original y su versión procesada, proporcionando una medida estándar para comparar la efectividad de algoritmos de procesamiento de imágenes. El PSNR se define como:

$$PSNR = 10 \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \quad [\text{dB}]$$

Donde MAX_I representa el valor máximo posible que puede tomar un píxel en la imagen, y MSE es el Error Cuadrático Medio (Mean Squared Error).

El PSNR se expresa en decibeles (dB), donde valores más altos indican mejor calidad de imagen al representar menor distorsión. Para imágenes de 8 bits, valores típicos oscilan entre 30-50 dB en aplicaciones de compresión, considerándose alta calidad cuando $PSNR > 35$ dB. En transmisión inalámbrica, valores aceptables se sitúan entre 20-25 dB (26). La métrica es particularmente útil para comparar diferentes algoritmos de procesamiento bajo condiciones controladas.

Structural Similarity Index Measure (SSIM):

Es una métrica perceptual de referencia completa diseñada para predecir la calidad percibida de imágenes y videos digitales mediante la evaluación de cambios en la información estructural. En el IEEE Transactions on Image Processing (27), SSIM constituye un modelo basado en percepción que incorpora fenómenos perceptuales importantes del sistema visual humano, incluyendo términos de enmascaramien-

to de luminancia y contraste. A diferencia de métricas tradicionales como MSE o PSNR que estiman errores absolutos, SSIM considera que los píxeles tienen fuertes interdependencias, especialmente cuando están espacialmente próximos, evaluando la degradación de imagen como un cambio percibido en la información estructural. El índice SSIM entre dos imágenes x e y se define como:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

Donde:

- μ_x y μ_y son las medias locales de x e y
- σ_x^2 y σ_y^2 son las varianzas locales de x e y
- σ_{xy} es la covarianza local entre x e y
- C_1 y C_2 son constantes de estabilización

Coefficiente de similitud de Dice (DSC):

Métrica de validación del índice de superposición espacial entre dos segmentaciones. El DSC es también un caso especial de la estadística kappa, que es usada para análisis de confiabilidad cuando hay un número mucho mayor de vóxeles de fondo que el de los vóxeles objetivo (28). El DSC viene descrito por la fórmula 4.1.

$$DSC(A, B) = \frac{2(A \cap B)}{A + B} \tag{4.1}$$

Intersección Sobre Unión (IoU):

Es una medida que se basa en el índice de Jaccard, el cual es un coeficiente que determina la similitud entre dos conjuntos de datos. En la detección de objetos, el IoU mide el área de superposición entre el cuadro delimitador previsto (B_p) y el cuadro delimitador de la verdad fundamental (B_g) dividido por el área de unión entre ellos (29). Lo anterior viene dado por la fórmula (29):

$$J(B_p, B_g) = IoU = \frac{area(B_p \cap B_g)}{area(B_p \cup B_g)} \tag{4.2}$$

Capítulo 5

Metodología

La metodología desarrollada se fundamenta en un enfoque experimental y comparativo, orientado a evaluar el desempeño de modelos de segmentación semántica profunda para la detección automática de tejido cancerígeno en imágenes histopatológicas de cáncer de mama. El diagrama de flujo de la metodología desarrollada se expone en la Figura 5.1.

5.1. Enfoque general

El enfoque adoptado fue cuantitativo y de tipo experimental, lo cual permitió establecer comparaciones objetivas entre múltiples modelos computacionales construidos bajo distintas configuraciones. Se priorizó un diseño iterativo y sistemático, donde cada variable (arquitectura, backbone y técnica de preprocesamiento) fue tratada como un componente crítico en la cadena de rendimiento. Esto responde a la naturaleza altamente sensible del diagnóstico histopatológico, donde incluso pequeñas variaciones en el procesamiento pueden traducirse en grandes diferencias clínicas.

Además, se optó por una estrategia comparativa, no solo para identificar la arquitectura más eficaz, sino también para comprender cómo interactúan los distintos componentes del pipeline, desde el filtrado de imágenes hasta la elección del codificador.

5.2. Conjunto de datos

La base de datos utilizada en este estudio se compone de imágenes histopatológicas de cáncer de mama triple negativo (TNBC), obtenidas a partir de Whole Slide Images (WSIs) teñidas con Hematoxilina y Eosina (H&E). Estas imágenes fueron seleccionadas cuidadosamente para representar regiones de interés (ROIs) con diversidad morfológica relevante para el análisis histopatológico.

En total, se extrajeron 161 ROIs, donde dos patólogos expertos proporcionaron etiquetas curadas para todas las ROIs, que se consideran como la verdad de referencia (ground truth).

Cada ROI fue dividida en parches de 512×512 píxeles, un tamaño adecuado para tareas de segmentación con redes convolucionales en contextos de imágenes de alta

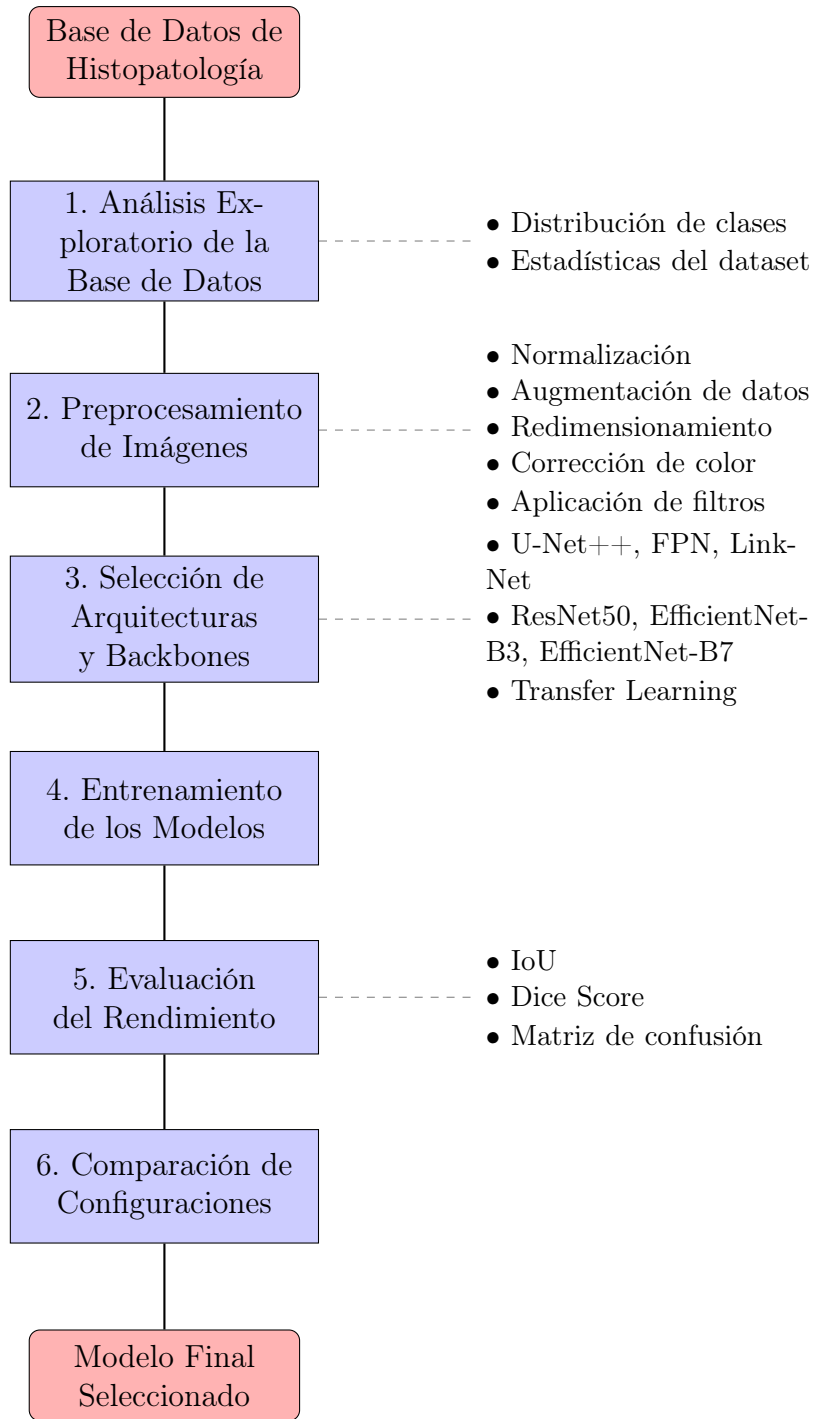


Figura 5.1: Metodología para Segmentación Semántica en Imágenes de Histopatología - Diagrama de flujo construcción del modelo.

resolución.

Conjunto	Número de parches
Entrenamiento	10,173
Validación	1,264
Prueba	399

Estas muestras fueron seleccionadas para incluir representaciones equilibradas desde el punto de vista espacial y contextual, aunque presentan una distribución desbalanceada de las clases anotadas.

El proceso de anotación se enfocó en identificar cinco clases tisulares principales:

- -1: Clase a ignorar, posiblemente usada para áreas que no deben ser evaluadas o segmentadas (bordes, artefactos, etc.).
- 0: Otros, representan áreas del tejido que no pertenecen a ninguna de las otras clases.
- 1: Tumor, probablemente la clase de mayor interés en la segmentación del tejido.
- 2: Estroma, tejido conectivo que rodea o sustenta el tumor.
- 3: Inflamación benigna, indicando áreas de inflamación que no están asociadas con el tumor.
- 4: Necrosis, áreas de tejido muerto, relevante para el análisis patológico.

Las imágenes estudiadas siguen la composición presentada en la Figura 5.2.

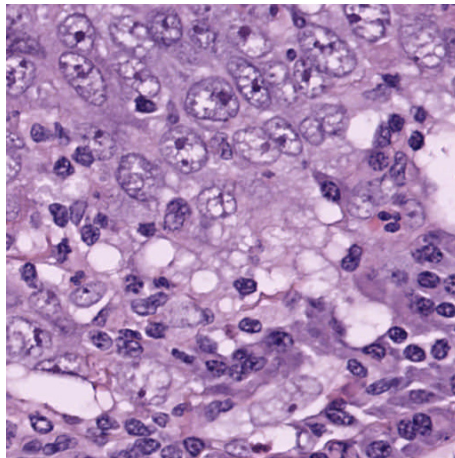


Figura 5.2: Imagen histopatológica de cáncer de mama.

En cuanto a la coloración, predominan tonos púrpuras y violetas debido a que se ha utilizado la técnica hematoxilina-eosina (H&E), la cual resalta núcleos y citoplasma. Los núcleos celulares aparecen más oscuros debido a la afinidad de la hematoxilina con los ácidos nucleicos. Luego, en las estructuras visibles se identifican núcleos celulares de forma ovalada o redonda distribuidos irregularmente en el campo. En la densidad celular se observa una alta densidad de núcleos en algunas áreas, con núcleos agrupados o superpuestos. Esto podría ser característico de un tejido proliferativo, como tejido tumoral o inflamado y por último, el fondo, la matriz extracelular, tiene una textura fibrosa, probablemente debido al estroma del tejido, que es visible entre los núcleos.

5.2.1. Distribución por clases

El dataset de histopatología de cáncer de mama triple negativo está compuesto por 6 clases distribuidas en tres conjuntos: entrenamiento (2.67M píxeles), validación (331K píxeles) y prueba (105K píxeles), totalizando más de 3.1 millones de píxeles anotados. El análisis de la distribución de clases revela un desbalance severo en el dataset, con dos clases dominantes (Tumor, Otra) que representan más del 70 % de los píxeles totales, mientras que clases minoritarias críticas para el diagnóstico (Necrosis) representa solo el 1,64 % del total de píxeles. La Tabla 5.1 muestra la distribución detallada por clase.

Clase	Entrenamiento Conteo	Validación Conteo	Prueba Conteo
Ignore	11.953.492 (0,4 %)	1.301.942 (0,4 %)	832.811 (0,8 %)
Other	112.034.021 (4,2 %)	13.670.605 (4,1 %)	8.081.445 (7,7 %)
Tumor	1.121.385.447 (42,0 %)	128.788.081 (38,87 %)	41.682.703 (39,9 %)
Stroma	912.564.269 (34,2 %)	114.775.521 (34,6 %)	35.727.408 (34,2 %)
Benign Inflammation	340.062.083 (12,8 %)	33.764.570 (10,2 %)	14.271.551 (13,6 %)
Necrosis	168.791.600 (6,3 %)	39.049.297 (11,8 %)	3.999.538 (3,8 %)
Total	2.666.790.912 (100 %)	331.350.016 (100 %)	104.595.456 (100 %)

Cuadro 5.1: Distribución de píxeles por clase

5.3. Preprocesamiento de imágenes

El preprocesamiento se diseñó no como un paso aislado, sino como una etapa estratégica para reducir el ruido, homogenizar la entrada al modelo y preservar características morfológicas clave del tejido. Por ello, se combinaron técnicas estándar, como redimensionamiento, normalización y conversión a tensores, con transformaciones específicas para aumentar la variabilidad (flip horizontal) y fomentar la generalización.

Adicionalmente, se exploró el impacto de filtros espaciales (Gaussiano, Bilateral, Wiener), los cuales se integraron como variables experimentales. Esta decisión metodológica responde a la necesidad de controlar el compromiso entre eliminación de ruido y pérdida de detalles estructurales. El filtro bilateral, por ejemplo, fue seleccionado como técnica definitiva por su capacidad para preservar bordes relevantes sin distorsionar la arquitectura del tejido, resultado que se sustenta tanto en métricas objetivas (PSNR, SSIM, Grad-SSIM) como en una evaluación visual cualitativa. El diagrama de flujo del pipeline del procesamiento de los datos se muestra en la Figura 5.3, en donde se describe el flujo detallado del pipeline de preprocesamiento.

to desarrollado. El proceso inicia con la carga de las imágenes en formato RGB, transformándolas desde su codificación original en BGR. Luego, se realiza un balance de color mediante la aplicación de la técnica CLAHE con un límite de recorte de 2.5, lo cual permite mejorar el contraste local sin sobreexponer regiones homogéneas. Esta operación se complementa con una corrección cromática del 20%, lo que ajusta la fidelidad del color a niveles estándar sin alterar la estructura del tejido.

Posteriormente, se ejecuta una etapa de detección automática de ruido con el fin de evaluar la calidad de la imagen. Si el nivel de ruido es superior a un umbral de 2.5, se aplica un filtrado avanzado que combina un filtro bilateral con un filtro de media no local (NLM), buscando preservar bordes y estructuras relevantes mientras se reduce el ruido. En caso contrario, se utiliza un filtrado más simple basado únicamente en el filtro bilateral, menos intensivo pero suficiente para imágenes con bajo nivel de ruido.

Una vez filtrada la imagen, se realiza una combinación entre la imagen original y la imagen filtrada, utilizando un factor de preservación del 70%, que permite retener parte de la información original para evitar una pérdida excesiva de detalles significativos. Después, se aplica una mejora de contraste compuesta por CLAHE, ajustes gamma y un factor de intensificación del 40%, contribuyendo a una visualización más clara de las estructuras celulares.

En la siguiente etapa, se implementa un realce de detalles utilizando filtros de Canny, con una intensidad moderada del 25%, lo que mejora los bordes finos y los contornos relevantes para facilitar la posterior segmentación. Finalmente, se realiza un procesamiento específico del tejido con un nivel de preservación del 60%, asegurando que la información estructural de las regiones de interés se mantenga en condiciones óptimas.

5.4. Arquitecturas y backbones

Se seleccionaron tres arquitecturas ampliamente validadas en segmentación médica: U-Net++, LinkNet y Feature Pyramid Network (FPN). Cada una presenta ventajas diferenciales: U-Net++ optimiza la transferencia de información mediante conexiones densas; LinkNet prioriza la eficiencia computacional y la reconstrucción rápida; FPN, por su parte, permite trabajar con información multiescala.

Estas arquitecturas se combinaron con tres codificadores preentrenados en ImageNet: EfficientNet-B3, EfficientNet-B7 y ResNet50. Esta decisión buscó evaluar cómo el tipo y profundidad del backbone influye en la extracción de características relevantes. EfficientNet-B7, por ejemplo, con su escalado compuesto, prometía un mejor reconocimiento de patrones finos, lo cual se confirmó en los resultados obtenidos.

5.5. Entrenamiento

Para el entrenamiento del modelo de segmentación semántica se utilizó la arquitectura LinkNet con codificador timm-efficientnet-b7 preentrenado en ImageNet, configurado para procesar imágenes RGB (3 canales) y clasificar píxeles en seis clases.

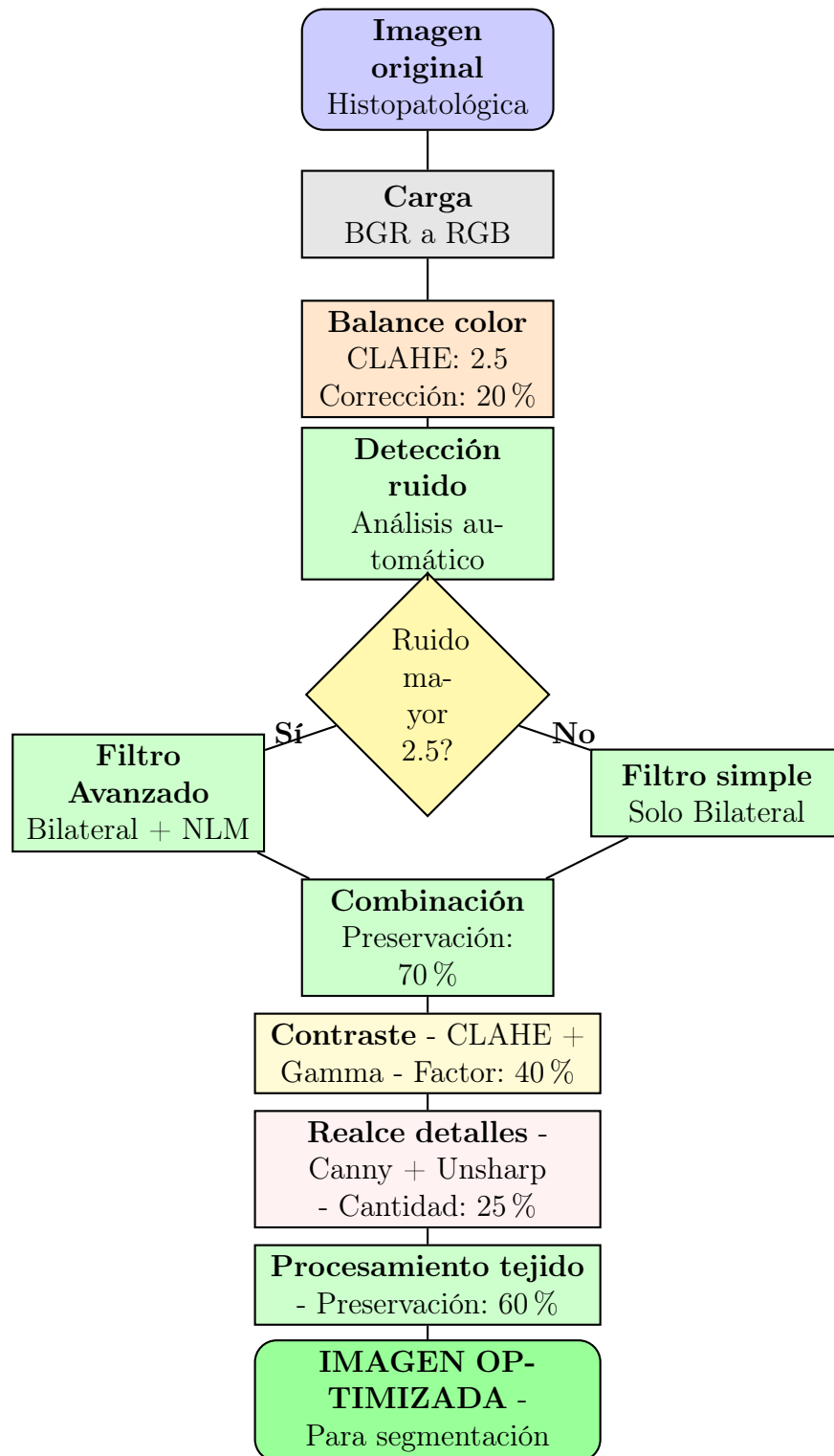


Figura 5.3: Pipeline de preprocesamiento adaptativo para imágenes histopatológicas.

Esta arquitectura fue seleccionada debido a su equilibrio entre precisión y eficiencia computacional, especialmente relevante en contextos clínicos donde la segmentación de tejido histológico requiere una alta sensibilidad sin comprometer el tiempo de inferencia.

El proceso de entrenamiento empleó 10 épocas, con un tamaño de lote ajustado previamente (`batch_size = 8`) y una tasa de aprendizaje fija (`learning_rate`) de 1×10^{-4} , utilizando el optimizador Adam. La elección del optimizador Adam obedece a su capacidad para adaptarse dinámicamente al gradiente y la escala del parámetro, lo que resulta en una convergencia más rápida y estable, especialmente en tareas con alto nivel de ruido y variabilidad como la segmentación médica. La tasa de aprendizaje se fijó en un valor conservador para evitar oscilaciones en la función de pérdida y asegurar una convergencia progresiva.

Respecto a la función de pérdida, se empleó la Dice Loss en modo multiclase y a partir de logits sin normalización (`from_logits=True`). Esta elección se justifica por su eficacia en tareas de segmentación con clases desbalanceadas, ya que la pérdida basada en el coeficiente de Dice penaliza directamente la discrepancia entre las máscaras predichas y reales, enfocándose en la superposición entre segmentos. Esta función también proporciona mayor sensibilidad para clases minoritarias, como ciertas estructuras histológicas en las imágenes utilizadas.

Una consideración crucial en la configuración del modelo fue la exclusión de la clase con etiqueta 5 (`ignore_index=5`), tanto en el entrenamiento como en el cálculo de métricas. Esta clase, denominada "Ignore", corresponde a regiones de la imagen marcadas como irrelevantes o inconsistentes para el análisis, ya sea por ambigüedad en la anotación, artefactos técnicos o regiones no representativas del tejido. Incluir estos píxeles en el proceso de entrenamiento podría inducir ruido en el aprendizaje del modelo y degradar su capacidad para distinguir patrones relevantes entre clases de interés. Por esta razón, se utilizó la opción `ignore_index=5` integrada a la función de pérdida (DiceLoss), excluyendo esta clase del conjunto de etiquetas consideradas para el cálculo de métricas de desempeño como Dice e IoU.

El uso de técnicas de aumento de datos fue clave para evitar sobreajuste, especialmente en un contexto con clases desbalanceadas como "Necrosis". El entrenamiento fue acelerado mediante GPU A100 en Google Colab Pro+, lo que permitió ejecutar múltiples modelos en paralelo y mantener consistencia en las condiciones experimentales.

5.6. Evaluación del rendimiento

El rendimiento de cada modelo se evaluó de forma exhaustiva, aplicando métricas de validación tanto globales como por clase. Se emplearon el Dice Score y el IoU, dos métricas ampliamente aceptadas en segmentación médica por su sensibilidad a la superposición espacial y su capacidad para capturar errores de clasificación.

Así mismo, se integraron representaciones visuales (matrices de confusión, curvas de pérdida y predicciones vs. ground truth) que complementaron las métricas numéri-

cas, permitiendo una interpretación cualitativa del comportamiento de cada modelo y facilitando el análisis clínico posterior.

5.7. Comparación de configuraciones

Finalmente, se construyó una tabla comparativa que resume los resultados de todas las configuraciones combinando arquitecturas, backbones y técnicas de filtrado. Esto permitió identificar las combinaciones más efectivas para la detección de tejido cancerígeno en cada clase, con especial énfasis en la clase de necrosis, por su relevancia clínica.

La comparación también permitió evaluar la influencia del preprocesamiento sobre la mejora del rendimiento del modelo en general y por clase. Se identificaron tendencias entre la arquitectura empleada, el tipo de backbone y el efecto del filtrado sobre la calidad de segmentación.

Esta evaluación integral proporciona evidencia para recomendar configuraciones óptimas en tareas futuras de segmentación de imágenes histopatológicas, contribuyendo al diagnóstico asistido por inteligencia artificial.

Capítulo 6

Resultados

6.1. Objetivo 1

Se evaluaron cuatro filtros: Gaussiano, Bilateral, Wiener y Difusión, comparando su desempeño en términos de calidad visual, fidelidad estructural y capacidad de reducción de ruido. Las métricas utilizadas para esta evaluación fueron PSNR, SSIM, Grad-SSIM y la variación del nivel de ruido estimado. Los resultados se exponen en la Tabla 6.1.

Filtro	PSNR	SSIM	Grad-SSIM	Red-ruido
Bilateral	31.16	0.8686	0.7212	0.0212
Wiener	25.68	0.8923	0.7721	-0.2629
Diffusion	14.04	0.2665	0.2885	-0.7625
Gaussian	34.42	0.9515	0.8576	0.0074

Cuadro 6.1: Comparación de métricas para elección de filtro.

El filtro Gaussiano presentó el desempeño más alto en términos de fidelidad global y estructural, con los valores máximos en PSNR, SSIM y Grad-SSIM. Las imágenes resultantes mostraron bordes suavizados de forma efectiva y una apariencia homogénea. No obstante, la reducción de ruido fue mínima (solo +0.0074), lo que sugiere que este filtro tiende a suavizar sin eliminar del todo el grano residual, además de que puede comprometer detalles finos importantes.

El filtro Bilateral ofreció un equilibrio más favorable entre fidelidad visual y limpieza. Alcanzó un PSNR razonable (31.16), un SSIM sólido (0.87) y un Grad-SSIM aceptable (0.72), junto con una leve pero positiva reducción de ruido (+0.0212). Aunque no logró los valores máximos en fidelidad global como el filtro Gaussiano, su capacidad de preservar bordes y estructuras relevantes fue superior, lo que lo convierte en una opción versátil para mantener la integridad del tejido histológico.

El filtro Wiener mostró un comportamiento mixto. Por un lado, alcanzó valores elevados de SSIM (0.8923) y Grad-SSIM (0.7721), lo que indica una buena preservación de bordes y estructuras. Sin embargo, su PSNR fue considerablemente más bajo (25.68), y se observó un incremento en el nivel de ruido (-0.2629), lo cual puede deberse a una amplificación de artefactos en el fondo de las imágenes.

Luego, el filtro de Difusion demostró un desempeño global deficiente. Presentó un PSNR extremadamente bajo (14.04), valores reducidos en SSIM y Grad-SSIM (ambos por debajo de 0.30), y un aumento significativo del ruido (-0.7625). En general, este filtro deterioró tanto la fidelidad como la estructura de las imágenes, intensificando el ruido en lugar de atenuarlo.

Aunque el filtro Gaussiano obtuvo los puntajes más altos en fidelidad visual, su efecto limitado en la reducción de ruido y su tendencia a suavizar excesivamente los detalles lo hacen menos adecuado para los fines del proyecto. El filtro Bilateral, en cambio, ofreció el mejor equilibrio entre limpieza, preservación estructural y nitidez, por lo que fue seleccionado como la técnica de preprocesamiento definitiva para el resto del flujo de trabajo.

Una vez seleccionada la mejor técnica de filtrado se desarrolló e implementó un pipeline de preprocesamiento adaptativo para imágenes histopatológicas de cáncer de mama, diseñado para preservar las características morfológicas del tejido mientras mejora la calidad de la imagen. Para evaluar la efectividad del preprocesamiento se usaron métricas estándar de calidad de imagen aplicadas al conjunto de imágenes de entrenamiento, como se observa en la Tabla 6.2.

Métrica	Valor promedio	Desviación estándar
PSNR	35.16 dB	± 1.08 dB
SSIM	0.9961	± 0.0009
Mejora de Nitidez	+4.7 %	± 2.7 %

Cuadro 6.2: Métricas de calidad del preprocesamiento.

Se encontró que el PSNR, que con un valor promedio de 35.16 dB, indica una buena calidad de preservación de la información original, demostrando que el procesamiento mantiene la integridad de los datos diagnósticos. En cuanto a la SSIM, el valor de 0.9961 (muy cercano al máximo teórico de 1.0) demuestra una buena preservación de las características estructurales de las imágenes. Finalmente, con la mejora de Nitidez, el incremento promedio del 4.7 % en la nitidez indica una mejora controlada en la definición de bordes y contornos celulares.

De igual manera, se observó que la baja desviación estándar en las métricas (PSNR: ± 1.08 dB, SSIM: ± 0.0009) demuestra la consistencia y robustez del algoritmo de preprocesamiento aplicado a diferentes tipos de imágenes histopatológicas dentro del dataset.

En la Figura 6.1, se observa la comparación visual del preprocesamiento donde se logra una optimización efectiva de las imágenes histopatológicas, preservando íntegramente las estructuras morfológicas del tejido cancerígeno mientras mejora significativamente la calidad técnica. Se observa además una normalización de la variabilidad cromática inherente a la tinción H&E, mayor definición en los contornos celulares y estructuras glandulares, y un contraste optimizado que facilita la diferenciación entre diferentes tipos de tejido.

Comparación Original vs Procesado (Modo Moderate)

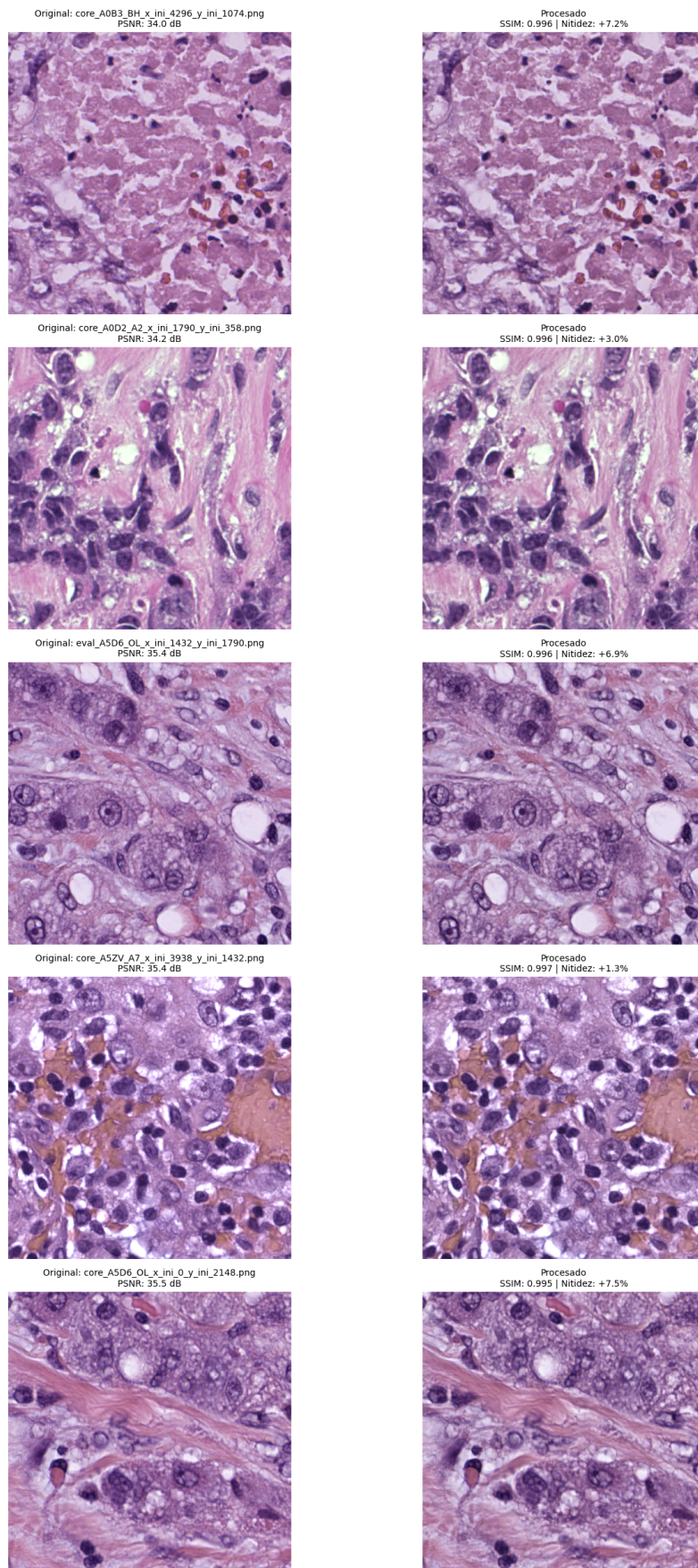


Figura 6.1: Comparación visual de preprocesamiento

6.2. Objetivo 2 y Objetivo 3

6.2.1. Procesamiento con imágenes crudas.

General:

Se evaluó el desempeño de los modelos de segmentación cuando fueron entrenados directamente sobre las imágenes crudas, es decir, sin aplicar ningún tipo de preprocesamiento previo. Para esta evaluación se utilizó el valor promedio del coeficiente de Dice como métrica de comparación entre las distintas combinaciones de arquitectura y backbone, como se observa en la Tabla 6.3.

Arquitectura	Backbone	Dice Promedio
LinkNet	EfficientNet-B7	0.7994
UNet++	EfficientNet-B7	0.7985
UNet++	EfficientNet-B3	0.7938
UNet++	ResNet50	0.7718
FPN	ResNet50	0.7619
FPN	EfficientNet-B7	0.7510
FPN	EfficientNet-B3	0.7444
LinkNet	ResNet50	0.6946
LinkNet	EfficientNet-B3	0.5488

Cuadro 6.3: Dice por arquitectura y *backbones* en imágenes crudas.

Entre los resultados, la arquitectura LinkNet con EfficientNet-B7 alcanzó el valor más alto de Dice promedio (0.7994), seguida muy de cerca por UNet++ con el mismo backbone (0.7985), concluyendo que ambas configuraciones mostraron un buen rendimiento sin preprocesamiento.

La arquitectura UNet++ también mostró buenos resultados al combinarse con EfficientNet-B3 (0.7938), así como con ResNet50 (0.7718), lo que evidencia un comportamiento estable en distintos escenarios. En cuanto a FPN, su mejor resultado se obtuvo con ResNet50 (0.7619), seguido por EfficientNet-B7 (0.7510) y EfficientNet-B3 (0.7444), presentando un descenso progresivo entre combinaciones.

LinkNet con ResNet50 y con EfficientNet-B3 registró los valores más bajos de Dice promedio, con 0.6946 y 0.5488, respectivamente, lo cual indica una menor capacidad de segmentación cuando se emplean estas combinaciones específicas sin preprocesamiento.

Por clase:

Se analizaron los valores del coeficiente de Dice por clase para cada combinación de arquitectura y backbone, utilizando imágenes crudas (sin preprocesamiento) de las clases distintas clases de tejido: “Other”, Tumor, Estroma, Inflamación Benigna y Necrosis, como se observa en la Tabla 6.4.

Arquitectura	Backbone	Other	Tumor	Stroma	Benign Inflamm.	Necrosis
LinkNet	EfficientNet-B7	0.8298	0.8387	0.7696	0.7533	0.8054
UNet++	EfficientNet-B7	0.8182	0.8342	0.7732	0.7709	0.7959
UNet++	EfficientNet-B3	0.8035	0.8407	0.7675	0.7767	0.7804
UNet++	ResNet50	0.7803	0.8306	0.7624	0.7462	0.7396
FPN	ResNet50	0.8154	0.8026	0.7436	0.7536	0.6943
FPN	EfficientNet-B7	0.7932	0.8353	0.7520	0.7265	0.6478
FPN	EfficientNet-B3	0.7793	0.8355	0.7541	0.7592	0.5940
LinkNet	ResNet50	0.4565	0.7831	0.7343	0.7663	0.7330
LinkNet	EfficientNet-B3	0.0209	0.8223	0.7563	0.7073	0.4373

Cuadro 6.4: Dice por clase por arquitectura y *backbones* en imágenes crudas.

La combinación de LinkNet con EfficientNet-B7 obtuvo los valores más altos en casi todas las clases, destacándose en Tumor (0.8387), "Other" (0.8298) y Necrosis (0.8054), con buen rendimiento también en Estroma (0.7696) e Inflamación Benigna (0.7533). La arquitectura UNet++, también con EfficientNet-B7, mostró un comportamiento similar, alcanzando 0.8342 en Tumor, 0.7732 en Estroma, 0.7709 en Inflamación Benigna y 0.7959 en Necrosis, manteniendo una buena consistencia entre clases.

En el caso de UNet++ con EfficientNet-B3, los valores fueron comparables: 0.8407 en Tumor, 0.7675 en Estroma y 0.7804 en Necrosis. Con ResNet50, UNet++ mantuvo un rendimiento estable, con valores entre 0.7462 y 0.8306 en las diferentes clases.

La arquitectura FPN con ResNet50 se caracterizó por una alta precisión en "Other" (0.8154) y desempeño sólido en Inflamación Benigna (0.7536) y Necrosis (0.6943). Con EfficientNet-B7 y EfficientNet-B3, FPN mostró valores similares en Tumor (0.8355) y Estroma (0.7520 – 0.7541), aunque con una reducción en Necrosis (0.6478 y 0.5940, respectivamente).

Por otro lado, las combinaciones de LinkNet con ResNet50 y especialmente con EfficientNet-B3 registraron un deterioro considerable en algunas clases. En particular, LinkNet–EfficientNet-B3 mostró un valor extremadamente bajo en "Other" (0.0209) y en Necrosis (0.4373), lo que sugiere una baja capacidad del modelo para generalizar sobre imágenes no preprocesadas en estas configuraciones. Incluso con ResNet50, LinkNet presentó caídas notables en "Other" (0.4565) y en Estroma (0.7343), a pesar de mantener valores aceptables en Tumor.

Evaluación

Observando las curvas de desempeño correspondientes a las imágenes crudas 1 2 3, se muestran que la mayoría de las arquitecturas lograron una convergencia estable durante el entrenamiento. En particular, LinkNet con EfficientNet-B7 presentó una mejora progresiva y sostenida tanto en la métrica Dice como en IoU, con una curva de pérdida decreciente y sin oscilaciones abruptas, lo que indica un buen ajuste del modelo al conjunto de datos sin sobreajuste evidente (Figura 6.2).

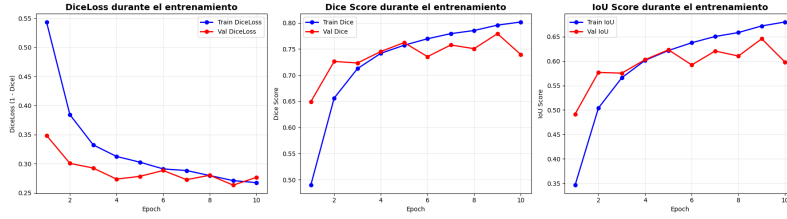


Figura 6.2: LinkNet-EfficientNet-B7 sin filtro

La arquitectura FPN con EfficientNet-B3 y LinkNet con EfficientNet-B7 también mostraron curvas suaves y consistentes, con estabilidad al final del entrenamiento, aunque con una pérdida residual ligeramente más alta que UNet++. En cambio, algunas combinaciones como LinkNet con ResNet50 evidenciaron mayor variabilidad en las curvas, lo que sugiere una menor capacidad de generalización o sensibilidad al tipo de entrada sin preprocesamiento.

Las curvas sin filtro reflejan que los modelos logran aprender representaciones útiles desde las imágenes crudas, pero con diferencias notables en la velocidad y estabilidad de convergencia según la arquitectura y el backbone utilizados.

Las matrices de confusión 4 5 6 indican que las clases Tumor y Estroma fueron las más consistentemente reconocidas por los modelos, con altos valores en la diagonal principal y baja confusión con otras clases.

Sin embargo, se observaron confusiones recurrentes entre la clase “Other” y Necrosis, especialmente en modelos como LinkNet con EfficientNet-B3, donde los errores de clasificación fueron más pronunciados. También se evidenció una tendencia a confundir Inflamación Benigna con Estroma en varias combinaciones, lo cual refleja posibles similitudes visuales no resueltas por los modelos sin asistencia de realce (Figura 6.3)

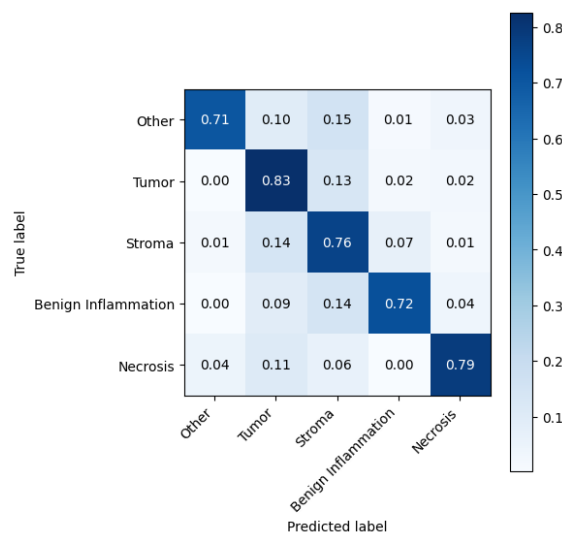


Figura 6.3: Matriz de confusión: LinkNet-EfficientNet-B7 sin filtro

La arquitectura UNet++, particularmente con EfficientNet-B7, logró una mejor se-

paración entre clases, mostrando una matriz más concentrada en la diagonal, mientras que FPN con ResNet50 presentó una distribución más dispersa, indicando confusión moderada en casi todas las clases.

6.2.2. Procesamiento con imágenes filtradas.

Se evaluaron distintas combinaciones de arquitecturas de segmentación semántica (LinkNet, FPN y UNet++) junto con tres tipos de backbones (EfficientNet-B7, EfficientNet-B3 y ResNet50), utilizando como métrica principal el coeficiente de similitud de Dice global. Estos resultados se observan en la Tabla 6.5.

Cuadro 6.5: Dice por arquitectura y *backbones* en imágenes filtradas.

Arquitectura	Backbone	Dice global
LinkNet	EfficientNet-B7	0.7971
LinkNet	EfficientNet-B3	0.7949
LinkNet	ResNet50	0.7882
FPN	ResNet50	0.7865
FPN	EfficientNet-B7	0.7816
FPN	EfficientNet-B3	0.7448
UNet++	EfficientNet-B7	0.7888
UNet++	EfficientNet-B3	0.7682
UNet++	ResNet50	0.7448

La combinación de LinkNet con EfficientNet-B7 obtuvo el valor de Dice global más alto (0.7971), destacándose por su buen desempeño en las clases de Tumor y Necrosis, ambas con valores superiores a 0.82. En segundo lugar, LinkNet con EfficientNet-B3 alcanzó un Dice global de 0.7949, mostrando una mayor capacidad de segmentación en la clase Necrosis (0.8331). Por su parte, LinkNet con ResNet50 obtuvo un Dice de 0.7882, con un rendimiento equilibrado y bajos niveles de solapamiento entre las clases Estroma y Tumor.

La arquitectura FPN con ResNet50 registró un Dice global de 0.7865, con un buen equilibrio en las clases Tumor (0.8111) y Estroma (0.7797). FPN con EfficientNet-B7 alcanzó un valor de 0.7816, mostrándose robusto en la clase Necrosis (0.7890), aunque presentó una caída en la clase Inflamación Benigna. Finalmente, FPN con EfficientNet-B3 obtuvo un Dice de 0.7448, con bajo rendimiento en Necrosis (0.6254) y menor consistencia general entre clases.

En cuanto a la arquitectura UNet++, la mejor combinación fue con EfficientNet-B7, logrando un Dice global de 0.7888, con un desempeño alto en la clase “Otros” (0.8395), aunque con mayor variabilidad entre clases. Con EfficientNet-B3, UNet++ alcanzó un valor de 0.7682, presentando una ligera caída en las clases Necrosis y Estroma. Por último, la combinación de UNet++ con ResNet50 obtuvo el menor Dice global (0.7448), mostrando un desempeño inferior y frecuentes solapamientos entre clases.

Por clase: Complementariamente, se calcularon los valores del coeficiente de Dice por clase para cada combinación de arquitectura y backbone, evaluando el desempe-

ño específico en las clases Other, Tumor, Estroma, Inflamación Benigna y Necrosis, los cuales se observan en la Tabla 6.6.

Arquitectura	Backbone	Other	Tumor	Stroma	Benign Inflamm.	Necrosis
LinkNet	EfficientNet-B7	0.8224	0.8323	0.7582	0.7472	0.8255
LinkNet	EfficientNet-B3	0.8177	0.8073	0.7524	0.7210	0.6254
LinkNet	ResNet50	0.7591	0.8259	0.7467	0.7751	0.7625
FPN	EfficientNet-B3	0.8300	0.8300	0.7677	0.7333	0.8151
FPN	EfficientNet-B7	0.7775	0.8315	0.7507	0.7592	0.7890
FPN	ResNet50	0.8361	0.8111	0.7797	0.7043	0.8106
UNet++	EfficientNet-B3	0.7407	0.8077	0.7453	0.7735	0.7736
UNet++	EfficientNet-B7	0.8395	0.8221	0.7479	0.7472	0.7871
UNet++	ResNet50	0.8177	0.8073	0.7524	0.7210	0.6254

Cuadro 6.6: Dice por clase por arquitectura y *backbones* en imágenes filtradas.

Dentro de la arquitectura LinkNet, la mejor combinación se observó con EfficientNet-B7, que alcanzó valores elevados y balanceados en todas las clases, destacándose en Tumor (0.8323) y Necrosis (0.8255). Esta arquitectura también mostró un rendimiento estable en Estroma (0.7582) y en la clase “Other” (0.8224). En contraste, al usar EfficientNet-B3, aunque se mantuvieron valores aceptables en Tumor (0.8073), se evidenció una caída considerable en Necrosis (0.6254), lo que sugiere una menor capacidad para segmentar tejido necrótico en esa configuración. Con ResNet50, LinkNet mantuvo un buen desempeño en Tumor (0.8259) e Inflamación Benigna (0.7751), aunque con menor consistencia en “Other” (0.7591).

En cuanto a FPN, la configuración con EfficientNet-B3 mostró un rendimiento sólido y homogéneo en todas las clases, especialmente en Necrosis (0.8151) y Tumor (0.83). Con EfficientNet-B7, los valores se mantuvieron en niveles similares, aunque con una leve disminución en Necrosis (0.789) y Estroma (0.7507). La combinación con ResNet50 presentó el valor más alto en Estroma (0.7797) y en “Other” (0.8361), pero una ligera caída en Inflamación Benigna (0.7043), indicando una posible sensibilidad a esta clase.

Para la arquitectura UNet++, la mejor combinación fue con EfficientNet-B7, que obtuvo el valor más alto en la clase “Other” (0.8395) y buenos niveles en Tumor (0.8221) y Necrosis (0.7871), aunque con variabilidad entre clases. Al emplear EfficientNet-B3, se evidenció un desempeño más bajo en general, particularmente en “Other” (0.7407), a pesar de mantener resultados aceptables en Tumor y Estroma. Finalmente, con ResNet50, UNet++ presentó una caída en Inflamación Benigna (0.721) y Necrosis (0.6254), replicando la baja observada con LinkNet-EfficientNet-B3 en esta última clase.

Evaluación

Adicionalmente, en las Figuras 7, 8, 9, se muestran la pérdida de todas las configuraciones presentan una reducción de la pérdida de entrenamiento durante las primeras 10 épocas, reflejando la capacidad de los modelos para aprender los patrones básicos del conjunto de datos. En particular, las arquitecturas con backbone EfficientNet-B7 alcanzan las menores pérdidas de validación y muestran una convergencia rápida y estable, con diferencias mínimas respecto a la pérdida de entrenamiento, lo que

sugiere un balance óptimo entre ajuste y generalización. En particular, con la arquitectura LinkNet (Figura 6.4). Por su parte, los modelos basados en EfficientNet-B3 logran resultados intermedios, con algo más de oscilación en la curva de validación tras la época 10, mientras que aquellos con ResNet-50 convergen más lentamente y se estabilizan en pérdidas de validación superiores.

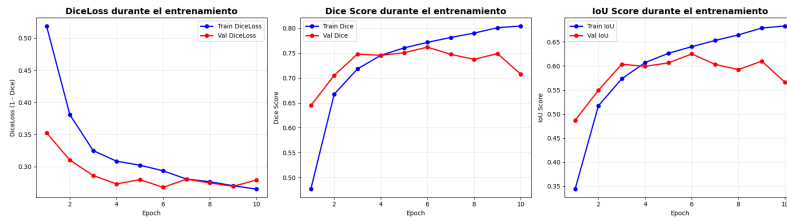


Figura 6.4: LinkNet-EfficientNet-B7 con filtro

Ahora, observando las matrices de confusión obtenidas tras aplicar el filtro 10 11 12, se muestran una mejor discriminación entre clases. En especial, las clases que anteriormente presentaban mayores niveles de confusión, como Necrosis e “Other”, ahora muestran una clasificación más precisa con mayor concentración en la diagonal. Particularmente, el rendimiento en la matriz de confusión de la arquitectura LinkNet con EfficientNet-B7 se muestra en la Figura 6.5.

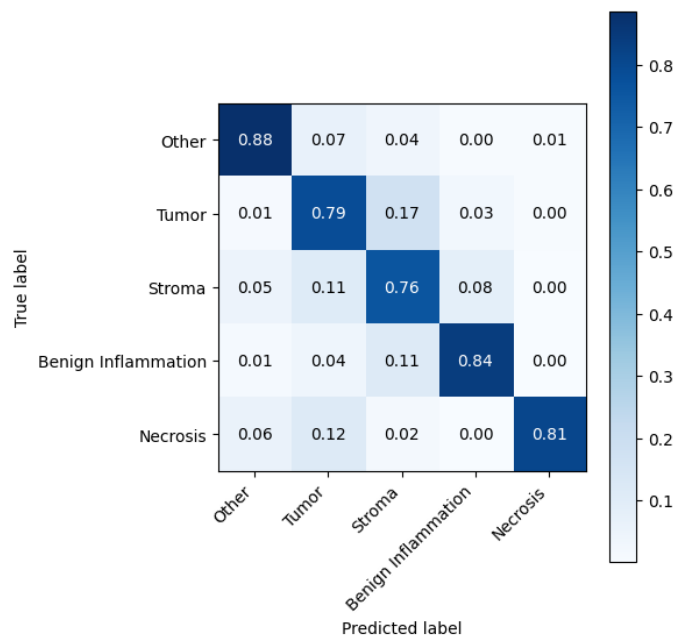


Figura 6.5: Matriz de confusión: LinkNet-EfficientNet-B7 con filtro

La arquitectura UNet++ con EfficientNet-B7 y LinkNet con EfficientNet-B7 presentan matrices de confusión más definidas, donde se reduce el solapamiento entre clases cercanas, lo que refleja una mejora directa en la capacidad del modelo para diferenciar los tipos de tejido. La clase In inflamación Benigna, que antes se confundía con Estroma, también fue mejor reconocida por modelos con entrada preprocesada.

En contraste, algunas configuraciones como FPN con ResNet50, aunque mostraron mejoras, mantuvieron cierto grado de dispersión, lo que sugiere que la efectividad del filtro depende tanto del preprocesamiento como de la arquitectura empleada.

Después de analizar las curvas de aprendizaje y las matrices de confusión, el modelo LinkNet con backbone EfficientNet-B7 se posiciona como la arquitectura óptima para la tarea de segmentación, alcanzando un Dice Score promedio de 0.7994. El comportamiento de las métricas durante el entrenamiento demuestra una convergencia estable y consistente: el DiceLoss desciende de manera suave desde 0.55 hasta 0.27, mientras que tanto el Dice Score como el IoU Score muestran una progresión ascendente consistente, alcanzando valores de convergencia de 0.80 y 0.67 respectivamente. Por otra parte, las curvas de entrenamiento y validación mantienen una proximidad notable a lo largo de las diez épocas, indicando que el modelo generaliza efectivamente sin mostrar signos de sobreajuste. Esta arquitectura combina la eficiencia del backbone EfficientNet-B7 para la extracción de características con las conexiones skip de LinkNet que preservan información espacial crítica, resultando en segmentaciones precisas y confiables. La estabilidad del proceso de entrenamiento, junto con las métricas finales obtenidas, confirman que esta configuración representa la mejor opción teniendo en cuenta que esta configuración no requiere preprocesamiento.

Finalmente, la capacidad del modelo para alcanzar un Dice Score cercano a 0.80 trabajando directamente con datos crudos representa una ventaja operacional considerable en entornos de producción donde la eficiencia computacional y la simplicidad del flujo de trabajo son prioritarias.

Las predicciones que se obtuvieron con la arquitectura Linknet con backbone EfficientNet-B7 entrenada con imágenes sin filtrar se expone en la Figura 6.6. En ella se puede evidenciar visualmente las similitudes entre las máscaras predichas y las verdaderas.

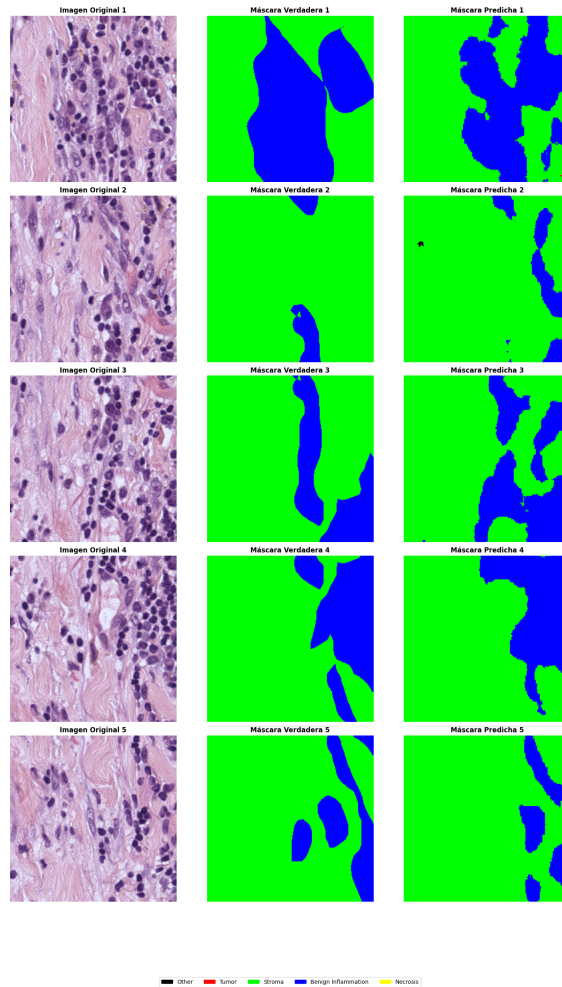


Figura 6.6: Predicciones LinkNet con EfficientNet-B7 entrenada con imágenes sin filtrar

Capítulo 7

Discusión

Los resultados en el presente trabajo evidencian que el uso de técnicas de preprocesamiento moderado, en conjunto con arquitecturas de segmentación semántica modernas como LinkNet, FPN y UNet++, logra una mejora consistente en la segmentación automática de tejido cancerígeno a partir de imágenes histopatológicas. Esta mejora se expresa tanto en métricas globales (Dice e IoU) como en métricas por clase, especialmente en tejidos clínicamente relevantes como Tumor y Necrosis.

En contraste, el enfoque desarrollado en este proyecto enfoque no se centró en la fuente de las anotaciones, sino en la calidad de la entrada visual, optimizando las imágenes antes del entrenamiento mediante filtrado bilateral y técnicas de mejora adaptativa. Esta diferencia metodológica llevó a que los modelos entrenados sobre imágenes filtradas alcanzaran valores de Dice superiores (hasta 0.7971 con LinkNet–EfficientNet-B7), superando incluso los valores reportados por el estudio basado en crowdsourcing. Este hallazgo sugiere que la mejora en la calidad de las imágenes puede ser tan determinante como el refinamiento de las etiquetas.

De igual manera, el trabajo de referencia identificó un patrón importante de confusión entre clases estructuralmente similares (por ejemplo, tejido tumoral vs. estroma), lo cual también se observó en nuestros resultados iniciales con imágenes crudas. Sin embargo, tras aplicar el preprocesamiento, las matrices de confusión mostraron una reducción clara en estos solapamientos, particularmente en modelos como UNet++ y LinkNet con EfficientNet-B7. Esto sugiere que el preprocesamiento no solo mejora la nitidez visual, sino que también puede facilitar la diferenciación morfológica por parte del modelo.

Por otra parte, ambos estudios coinciden en que la clase Necrosis representa un desafío considerable. En el artículo citado, esta clase también generó errores frecuentes, mientras que en el presente trabajo solo las combinaciones más robustas (FPN–EfficientNet-B3 o LinkNet–EfficientNet-B7) lograron superar valores de Dice del 0.80 para esta clase. Esto resalta la necesidad compartida de desarrollar estrategias especializadas para clases minoritarias, como aumentación dirigida o pérdidas adaptativas.

En conjunto, esta comparación respalda que, si bien la calidad de las etiquetas es crucial, la calidad visual del dato de entrada y la arquitectura del modelo juegan un

rol igual o incluso más determinante en el rendimiento final. El presente proyecto propone, por tanto, una ruta alternativa y complementaria a la del crowdsourcing, enfocada en optimizar el pipeline desde la imagen hacia el modelo, sin depender exclusivamente del desarrollo manual.

Capítulo 8

Conclusiones

En relación con el primer objetivo específico, que consistió en realizar el preprocesamiento de imágenes histopatológicas de cáncer de mama, se concluye que la implementación de un pipeline adaptativo, basado en la combinación de técnicas de corrección cromática, mejora de contraste y aplicación de filtros espaciales, permitió optimizar significativamente la calidad visual de las imágenes. La selección del filtro bilateral como técnica definitiva fue respaldada tanto por métricas objetivas (PSNR de 35.16 dB, SSIM de 0.9961) como por evaluaciones cualitativas, evidenciando su capacidad para reducir el ruido preservando al mismo tiempo las estructuras morfológicas clave del tejido. Por lo tanto, puede decirse que resultados encontrados reflejan que el preprocesamiento no solo mejora la calidad visual de las imágenes, sino que constituye una etapa fundamental para garantizar la integridad de la información diagnóstica.

Respecto al segundo objetivo, el cual buscaba desarrollar y entrenar modelos de segmentación semántica basados en técnicas de aprendizaje profundo, se concluye que las arquitecturas seleccionadas (UNet++, FPN y LinkNet), combinadas con backbones eficientes como EfficientNet-B3, EfficientNet-B7 y ResNet50, presentaron un desempeño robusto en la segmentación de tejido cancerígeno. Entre todas las combinaciones, destacó la arquitectura LinkNet con el backbone EfficientNet-B7, tanto con imágenes crudas como filtradas, mostrando una convergencia estable y métricas de desempeño consistentes a lo largo del entrenamiento, siendo la mejor configuración que pudo aprender mejor las características del dataset estudiado. Este modelo se consolidó como una alternativa eficiente y precisa, capaz de aprender representaciones discriminativas incluso sin preprocesamiento, lo que sugiere su relevancia para aplicaciones en entornos clínicos donde la eficiencia computacional es prioritaria.

Finalmente, en relación con el tercer objetivo, que consistió en evaluar y comparar el rendimiento de los modelos desarrollados, se concluye que el preprocesamiento contribuye significativamente a mejorar la capacidad de los modelos para distinguir entre clases tisulares similares. Esto se evidenció en la mejora de las matrices de confusión y el aumento del coeficiente de Dice por clase, particularmente en clases difíciles como Necrosis e Inflamación Benigna. Las configuraciones con preprocesamiento alcanzaron valores de Dice superiores al 0.79, mientras que las mejores configuraciones sin preprocesamiento también lograron métricas competitivas cercanas al 0.80. Esta evaluación comparativa permitió no solo seleccionar el modelo

óptimo, sino también identificar las condiciones experimentales bajo las cuales se maximiza su rendimiento, aportando así evidencia para futuras implementaciones en contextos de diagnóstico asistido por inteligencia artificial.

En conjunto, los hallazgos de este proyecto reafirman el potencial del aprendizaje profundo como herramienta de apoyo en el diagnóstico histopatológico del cáncer de mama. La integración de un preprocesamiento estratégico con arquitecturas modernas de segmentación semántica permitió obtener resultados precisos, consistentes y clínicamente relevantes. Estos resultados no solo responden satisfactoriamente a los objetivos planteados, sino que también siembra la posibilidad de futuras investigaciones orientadas a la implementación de estos modelos en entornos reales, donde la rapidez, la precisión y la automatización del análisis son fundamentales para mejorar la atención en salud.

Capítulo 9

Trabajos futuros

Como trabajos futuros en segmentación semántica, se pueden explorar nuevas arquitecturas de redes neuronales que mejoren la precisión y eficiencia de los modelos, así como el desarrollo de enfoques híbridos que combinen distintas técnicas de segmentación. También es importante avanzar en métodos de aprendizaje semi-supervisado y auto-supervisado para reducir la dependencia de grandes cantidades de datos etiquetados. Otra línea relevante es la integración multimodal, combinando diferentes tipos de datos para enriquecer la segmentación. Además, la optimización de modelos para su uso en tiempo real y en dispositivos con recursos limitados abre grandes oportunidades.

El manejo avanzado del desbalance de clases y la propuesta de métricas específicas más robustas son también áreas que requieren mayor atención. Finalmente, optimizar los procesos de preprocesamiento y postprocesamiento permitirá obtener resultados más fiables y precisos, facilitando la integración de estos modelos en sistemas automatizados de diagnóstico y control.

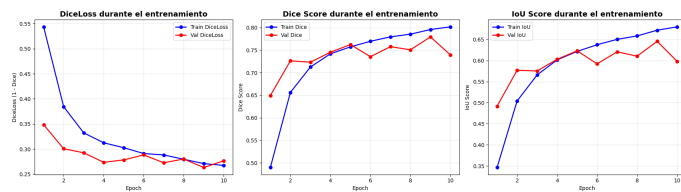
Bibliografía

- [1] S. F, S. A, and D. P, “Advancements in image classification using convolutional neural network,” in *2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*. IEEE, 2018.
- [2] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, ser. Lecture Notes in Computer Science, vol. 9351. Springer, Cham, 2015, pp. 234–241.
- [3] A. Chaurasia and E. Culurciello, “Linknet: Exploiting encoder representations for efficient semantic segmentation,” arXiv preprint arXiv:1707.03718, 2017. [Online]. Available: <https://arxiv.org/abs/1707.03718>
- [4] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 936–944.
- [5] D. S. C *et al.*, “Breast cancer statistics, 2019,” *CA: A Cancer Journal for Clinicians*, vol. 69, pp. 438–451, 2019.
- [6] S. S and G. M, “Machine learning methods for computer-aided breast cancer diagnosis using histopathology: A narrative review,” *Journal of Medical Imaging and Radiation Sciences*, pp. 1–12, 2019.
- [7] L.-P. M, M. Álvarez P, C. LAD, M. R, and K. AK, “Crowdsourcing segmentation of histopathological images using annotations provided by medical students,” in *Artificial Intelligence in Medicine*. Springer Nature Switzerland, 2023, pp. 245–249.
- [8] R. KKD, K. G, S. K, D. D, and R. S, “A review of medical image segmentation algorithms,” *EAI Endorsed Transactions on Pervasive Health Technologies*, p. 169184, July 2018.
- [9] S. K *et al.*, “Semantic segmentation of pancreatic cancer in endoscopic ultrasound images using deep learning approach,” *Cancers*, vol. 14, no. 20, p. 5111, 2022.
- [10] H. Y, C. W, H. AA, C. H, and Z. X, “A solution to the stagnation of multi-verse optimization: An efficient method for breast cancer pathologic images segmentation,” *Biomedical Signal Processing and Control*, vol. 86, p. 105208, 2023.

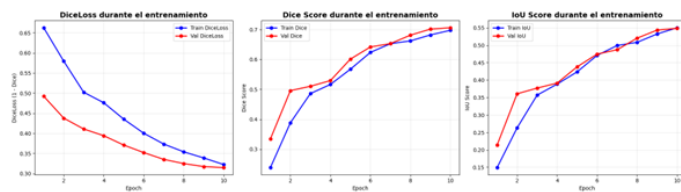
-
- [11] D. WR, B. N, and S. L, “Breast cancer histopathological images segmentation using deep learning,” *Sensors*, vol. 23, no. 17, p. 7318, 2023.
- [12] A. JS *et al.*, “Artificial intelligence in breast cancer diagnosis and personalized medicine,” *Journal of Breast Cancer*, vol. 26, 2023.
- [13] D. S, *Hands-On Image Processing with Python: Expert techniques for advanced image analysis and effective interpretation of image data*. Packt Publishing, 2018.
- [14] J. Chaki and N. Dey, *A Beginner’s Guide to Image Preprocessing Techniques*. CRC Press, 2018.
- [15] S. M. Florez-Aroni, M. A. Hanco-Condori, and F. Torres-Cruz, “Noise reduction in medical images,” *arXiv preprint arXiv:2301.01437*, 2023. [Online]. Available: <https://arxiv.org/abs/2301.01437>
- [16] M. Santamaría and B. Aldalur, “Realce de imágenes: filtrado espacial,” *Revista de Teledetección*, no. 17, pp. 31–42, 2002. [Online]. Available: <https://www.aet.org.es/revistas/revista17/AET17-04.pdf>
- [17] A. Buades, B. Coll, and J. M. Morel, “Non-local means denoising,” *Image Processing On Line*, vol. 1, pp. 208–212, 2011. [Online]. Available: https://doi.org/10.5201/ipol.2011.bcm_nlm
- [18] J. Canny, “A computational approach to edge detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, 1986. [Online]. Available: <https://doi.org/10.1109/TPAMI.1986.4767851>
- [19] G. M. H. Amer, M. A. Sid-Ahmed, and A. A. Farag, “On the canny edge detector,” *Pattern Recognition*, vol. 33, no. 5, pp. 853–867, 2000. [Online]. Available: [https://doi.org/10.1016/S0031-3203\(99\)00146-7](https://doi.org/10.1016/S0031-3203(99)00146-7)
- [20] B. CM and B. H, *Deep Learning*. Springer International Publishing, 2024.
- [21] K. JD, *Deep Learning*. MIT Press, 2019.
- [22] B. VE, K. R, and S. R, Eds., *Recent Trends and Advances in Artificial Intelligence and Internet of Things*. Springer International Publishing, 2020.
- [23] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, ser. Lecture Notes in Computer Science, vol. 9351. Springer, 2015, pp. 234–241. [Online]. Available: <https://arxiv.org/abs/1505.04597>
- [24] M. Tan and Q. V. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 2019, arXiv:1905.11946. [Online]. Available: <https://arxiv.org/abs/1905.11946>
- [25] ITU-T, “Reference algorithm for computing peak signal to noise ratio of a processed video sequence,” International Telecommunication Union, ITU-T Recommendation J.340, 2010.

- [26] ScienceDirect, “Peak signal-to-noise ratio,” Computer Science Topics, Elsevier, 2022. [Online]. Available: <https://www.sciencedirect.com/topics/computer-science/peak-signal-to-noise-ratio>
- [27] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004. [Online]. Available: <https://doi.org/10.1109/TIP.2003.819861>
- [28] Z. KH *et al.*, “Statistical validation of image segmentation quality based on a spatial overlap index,” *Academic Radiology*, vol. 11, no. 2, pp. 178–189, 2004.
- [29] P. R, N. SL, and da Silva EAB, “A survey on performance metrics for object-detection algorithms,” in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. IEEE, 2020.

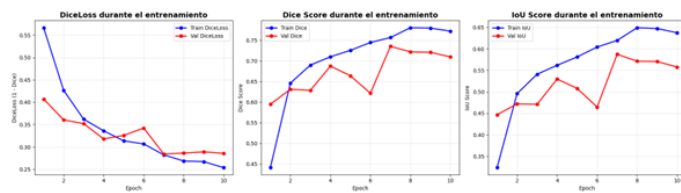
Anexos



1: LinkNet-EfficientNet-B7

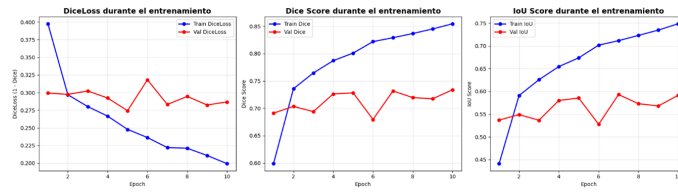


2: LinkNet-EfficientNet-B3



3: LinkNet-ResNet50

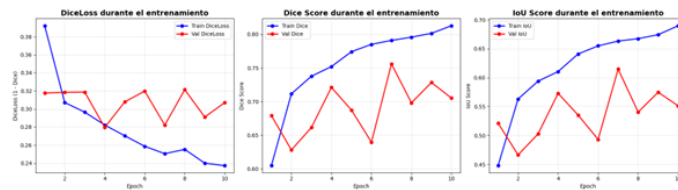
Figura 1: Arquitectura LinkNet con imágenes sin filtro



1: FPN-EfficientNet-B7

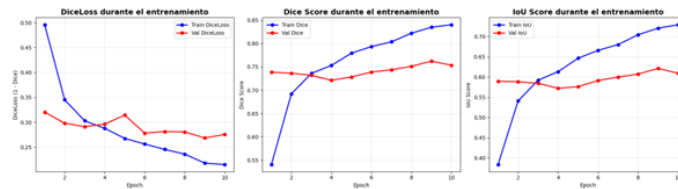


2: FPN-EfficientNet-B3



3: FPN-ResNet50

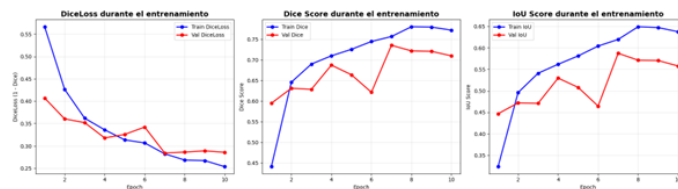
Figura 2: Arquitectura FPN con imágenes sin filtro



1: UNet++ EfficientNet-B7

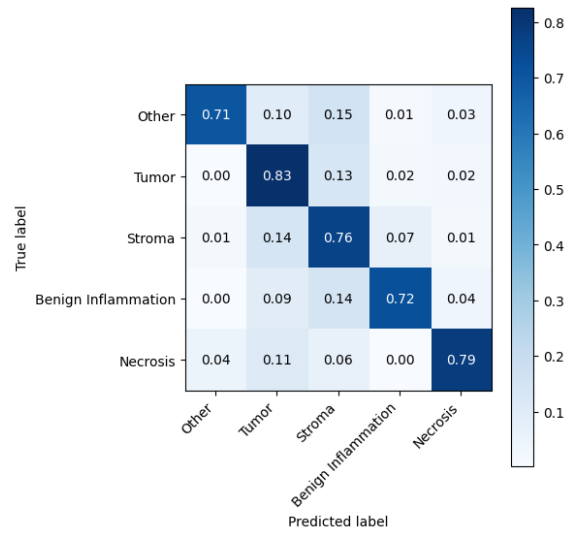


2: UNet++ EfficientNet-B3

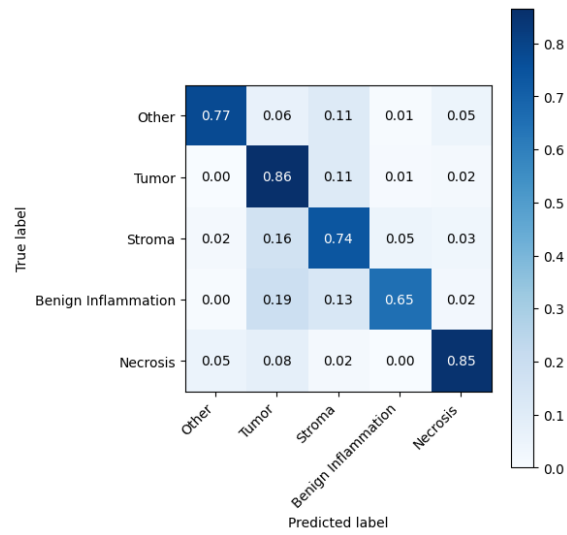


3: UNet++ ResNet50

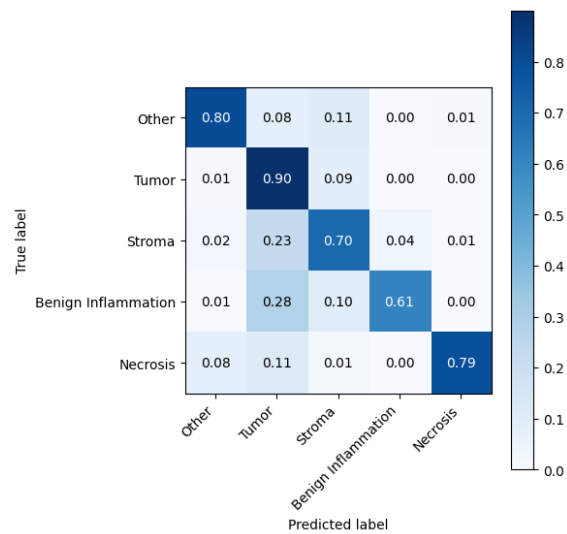
Figura 3: Arquitectura UNet++ con imágenes sin filtro



1: LinkNet-Efficient Net-B7

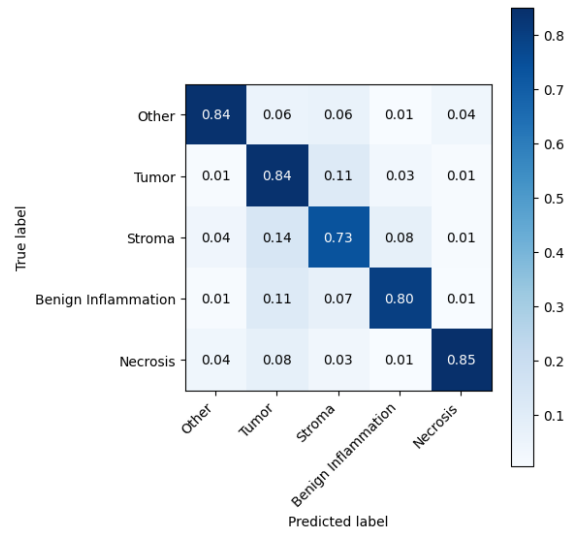


2: LinkNet-Efficient Net-B3

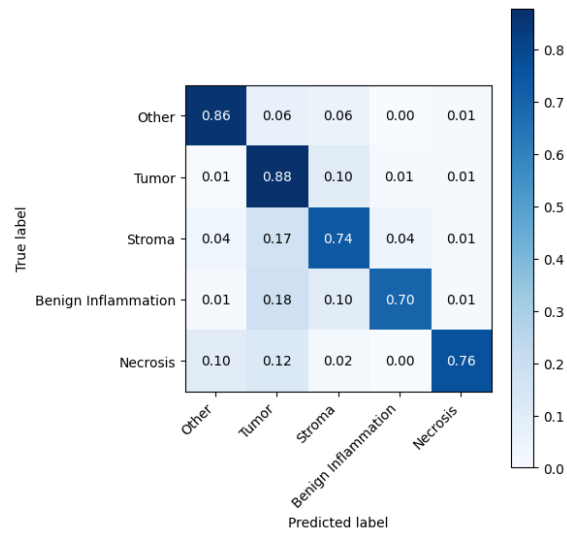


3: LinkNet-Efficient Net-ResNet50

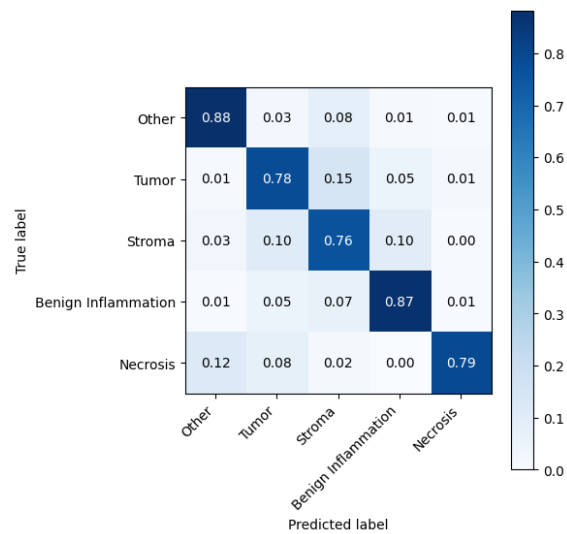
Figura 4: Arquitectura LinkNet con imágenes filtradas



1: FPN-Efficient Net-B7

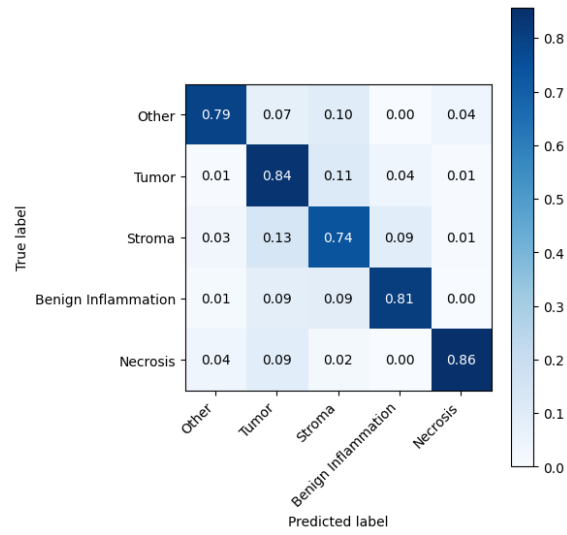


2: FPN-Efficient Net-B3

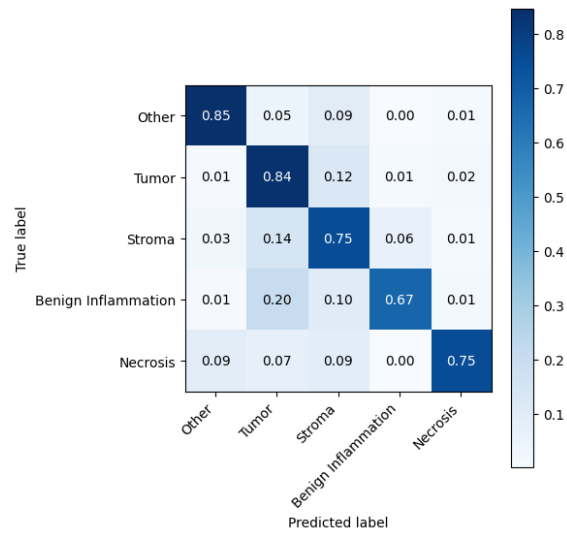


3: FPN-Efficient Net-ResNet50

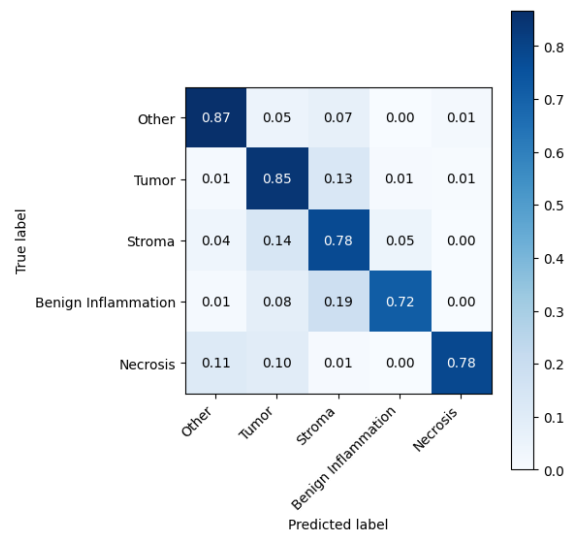
Figura 5: Arquitectura FPN con imágenes filtradas



1: UNet++-Efficient Net-B7

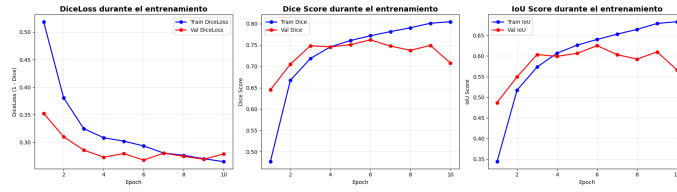


2: UNet++-Efficient Net-B3

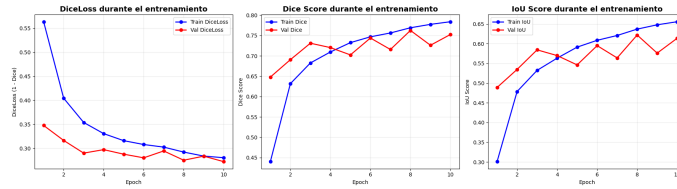


3: UNet++-Efficient Net-ResNet50

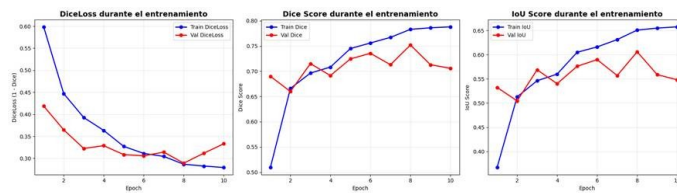
Figura 6: Arquitectura UNet++ con imágenes filtradas



1: LinkNet-Efficient Net-B7

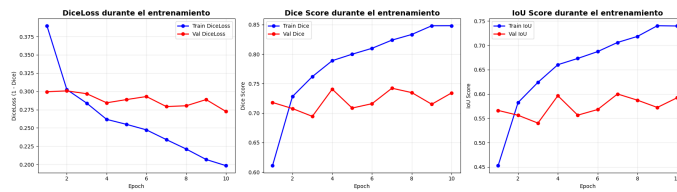


2: LinkNet-Efficient Net-B3

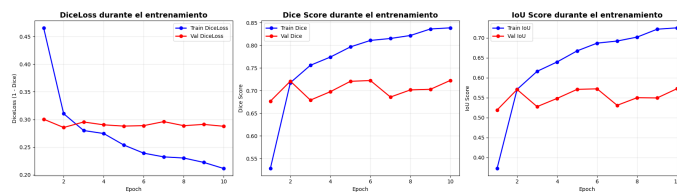


3: LinkNet-Efficient Net-ResNet50

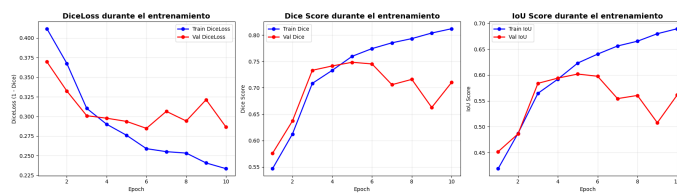
Figura 7: Arquitectura LinkNet con imágenes filtradas



1: FPN-Efficient Net-B7

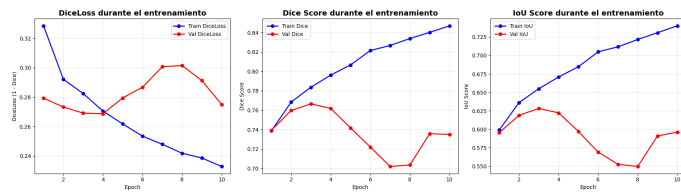


2: FPN-Efficient Net-B3

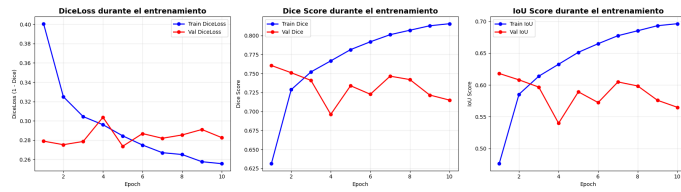


3: FPN-Efficient Net-ResNet50

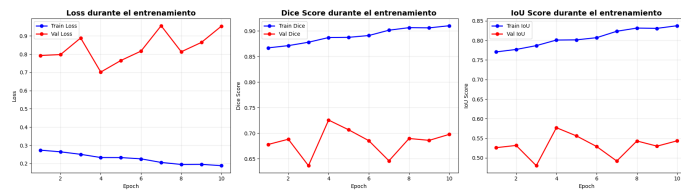
Figura 8: Arquitectura FPN con imágenes filtradas



1: UNet++-Efficient Net-B7

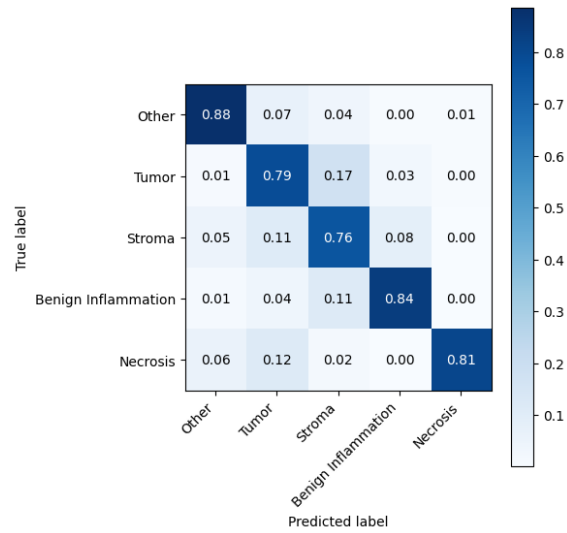


2: UNet++-Efficient Net-B3

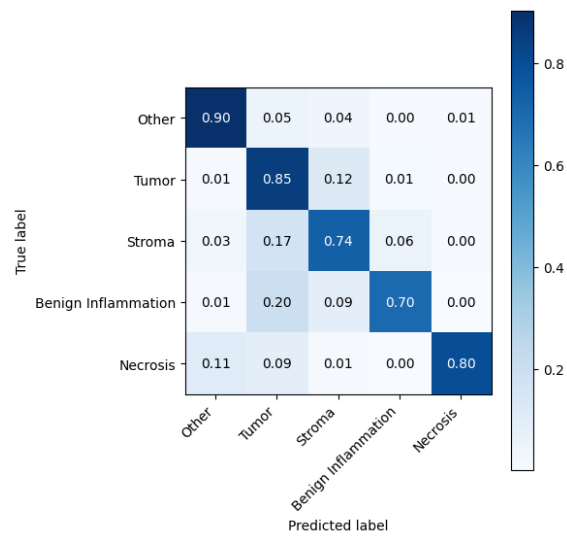


3: UNet++-Efficient Net-ResNet50

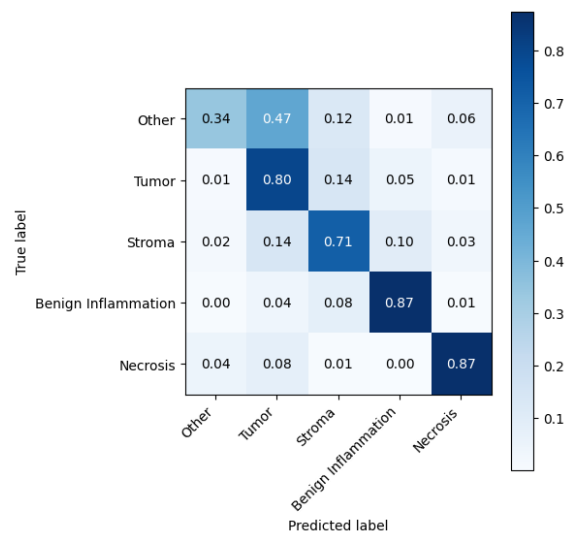
Figura 9: Arquitectura UNet++ con imágenes filtradas



1: LinkNet-Efficient Net-B7

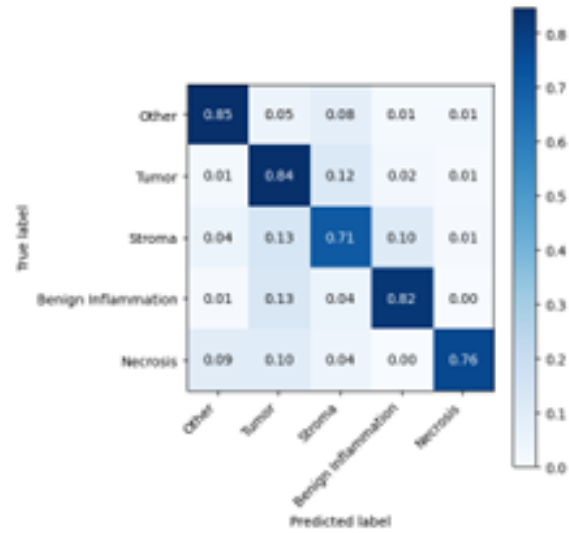


2: LinkNet-Efficient Net-B3

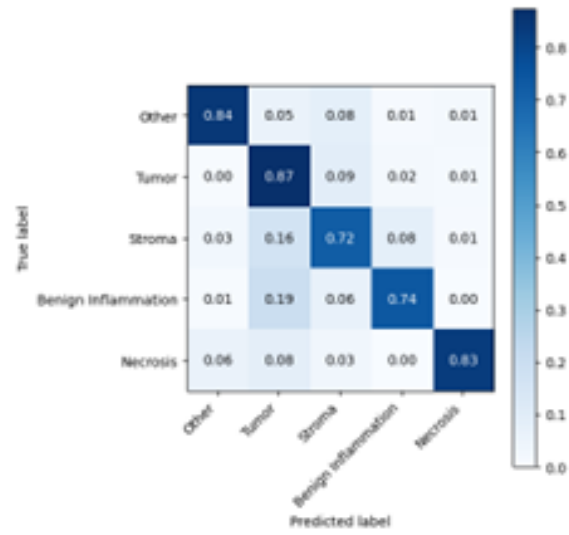


3: LinkNet-Efficient Net-ResNet50

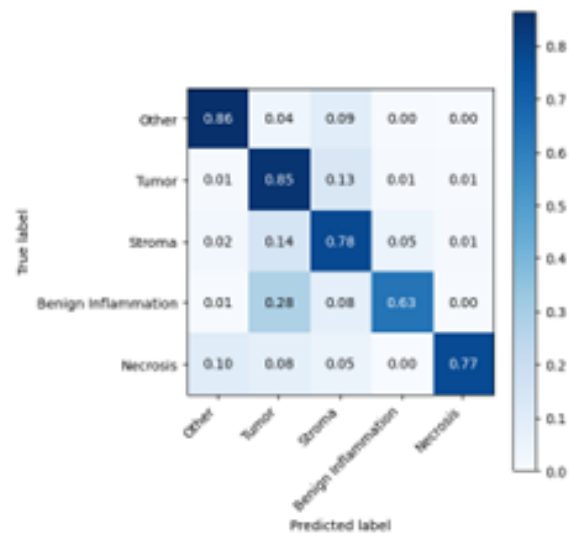
Figura 10: Arquitectura LinkNet con imágenes filtradas



1: FPN-Efficient Net-B7

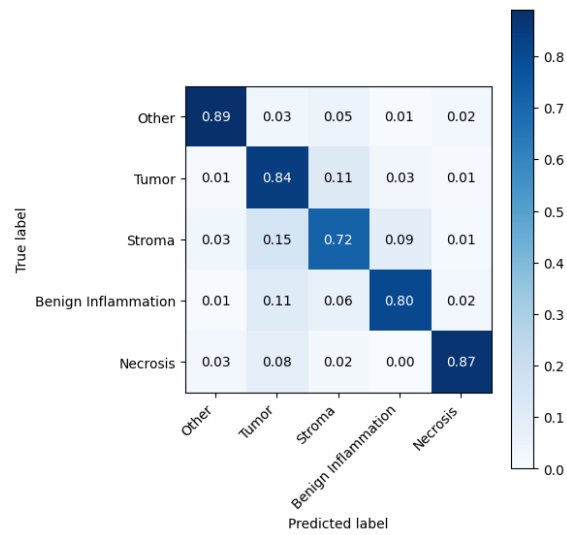


2: FPN-Efficient Net-B3

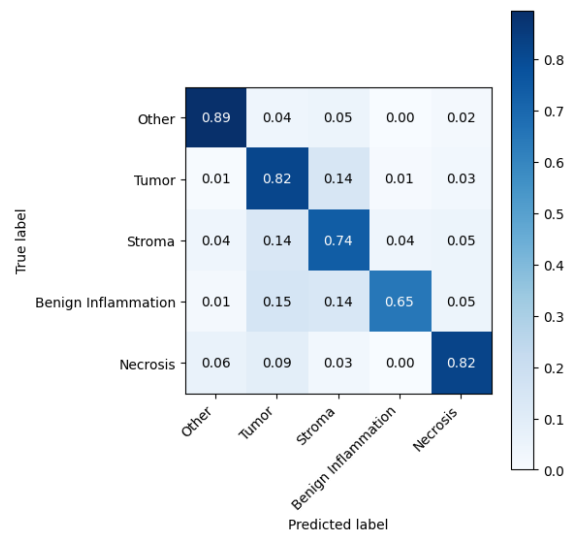


3: FPN-Efficient Net-ResNet50

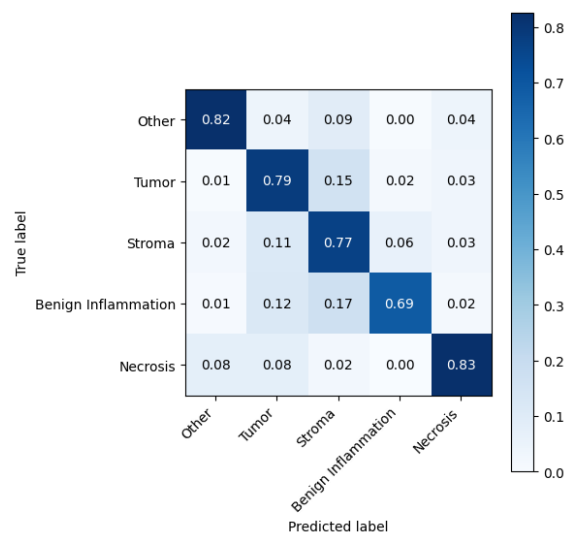
Figura 11: Arquitectura FPN con imágenes filtradas



1: UNet++-Efficient Net-B7



2: UNet++-Efficient Net-B3



3: UNet++-Efficient Net-ResNet50

Figura 12: Arquitectura UNet++ con imágenes filtradas