



**Maestría en Ciencia de Datos
Facultad de Ingeniería y Ciencias**

**FICHA RESUMEN
TRABAJO DE GRADO DE MAESTRÍA**

TITULO: “COMPARACIÓN DE MODELOS DE SEGMENTACIÓN DE ÁREA METASTÁSICA EN IMÁGENES HISTOPATOLÓGICAS DE CÁNCER DE MAMA PARA ESTIMAR CARGA TUMORAL”

1. ÉNFASIS: Sistemas y Computación
2. TIPO DE PROYECTO: Aplicado
3. ÁREA DE TRABAJO: Modelos computacionales
4. ESTUDIANTE: Diego Andrés Castellano Carvajal
5. CORREO ELECTRÓNICO: diegocastellano@javerianacali.edu.co
6. DIRECCIÓN Y TELÉFONO: Av 7 A # 4-45 Barrio San Luis, Cúcuta, Norte de Santander, 3105895877
7. DIRECTOR: Sergio Alexander Castro Casadiego
8. VINCULACIÓN DEL DIRECTOR (en la universidad): Externo
9. CORREO ELECTRÓNICO DEL DIRECTOR: sergio.castroc@ufps.edu.co
10. CO-DIRECTOR: Carlos Vicente Niño Rondón
11. GRUPO O EMPRESA QUE LO AVALA (Si aplica): Universidad de Pamplona
12. OTROS GRUPOS O EMPRESAS: Grupo de Investigación y Desarrollo en Electrónica y Telecomunicaciones, GIDET, Universidad Francisco de Paula Santander
13. PALABRAS CLAVE: Cáncer de mama, Ganglio linfático, Metástasis, Segmentación, Modelos computacionales
14. ODS QUE APLICA EL PROYECTO (Agenda 2030): ODS 9 y ODS 11
15. FECHA DE INICIO (Desarrollo del proyecto): 12/12/2024
16. RESUMEN: El cáncer de mama se considera la neoplastia maligna más común en mujeres. En 2022, según los reportes de Organización Mundial de la Salud 2.3 millones de mujeres fueron diagnosticadas con cáncer de mama en todo el mundo y 670.000 murieron. En el proceso de crecimiento del cáncer de mama es probable que se produzca metastáticos en los ganglios linfáticos en la medida que el tumor se propague en los tejidos mamaros circundantes. El análisis manual de muestras histopatológicas para el reconocimiento de regiones metastásicas es una tarea demandante de tiempo y es propensa a la variabilidad entre los patólogos. En el presente trabajo, se muestra la implementación y comparación de modelos de aprendizaje profundo para segmentar regiones metastásicas y estimar carga tumoral. Se abordó la alta heterogeneidad intermuestral y la variabilidad inherente al análisis manual de muestras histopatológicas. Se utilizó el conjunto de datos CAMELYON16, aplicando técnicas de preprocesamiento como la normalización de color y el aumento de datos para mejorar la generalización. Se compararon cuatro arquitecturas, incluyendo el modelo híbrido propuesto UNet + ResNet50, Junto con ResNet50, UNet y DenseNet121. Los resultados establecieron que el modelo híbrido fue la solución más robusta y equilibrada, superando consistentemente a los demás modelos que presentaron signos de sobreajuste. El modelo híbrido demostró la mayor precisión en la estimación de la carga tumoral, manteniendo su superioridad integral y un equilibrio en las métricas clave.



Pontificia Universidad
JAVERIANA
Cali

**COMPARACIÓN DE MODELOS DE SEGMENTACIÓN DE ÁREA METASTÁSICA EN
IMÁGENES HISTOPATOLÓGICAS DE CÁNCER DE MAMA PARA ESTIMAR CARGA
TUMORAL**

Diego Andrés Castellano Carvajal
Código 9014670

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

Director
PhD. Sergio Alexander Castro Casadiego

Codirector
MSc. Carlos Vicente Niño Rondón

FACULTAD DE INGENIERÍA Y CIENCIAS
MAESTRÍA EN CIENCIA DE DATOS
SANTIAGO DE CALI, ENERO DE 2026

AGRADECIMIENTOS

Agradezco infinitamente a Dios y al Espíritu Santo por la sabiduría recibida para lograr mis propósitos y metas. A mi mamá María Zulay y abuela Flor de María, quienes día a día me apoyan y me acompañan en cada una de mis metas. A mi familia, quienes me desean lo mejor en cada etapa de mi vida.

Agradezco al ingeniero Sergio Castro, director del proyecto, por la disposición y asesoramiento durante el desarrollo del proyecto aplicado. También, agradezco enormemente a mi codirector, el ingeniero Carlos Vicente Niño, quien agradezco infinitamente por su apoyo desde el pregrado y considero un gran amigo. Al ingeniero Byron Medina, por su apoyo y contribución en mi formación profesional.

DEDICATORIA

Dedico este logro a mi mamá, María Zulay, y a mi abuela Flor de María, quienes amo con todo mi corazón y son mi motor de vida. A mi tía Myriam y madrina Ana Milena, quienes me brindan su cariño. A mis primos Santiago, Valeria y Edwin David, quienes los considero mis hermanos.

TABLA DE CONTENIDO

INTRODUCCIÓN	7
1. DEFINICIÓN DEL PROBLEMA.....	9
1.1. PLANTEAMIENTO DEL PROBLEMA	9
1.2. FORMULACIÓN DEL PROBLEMA	10
2. OBJETIVOS DEL PROYECTO	11
2.1. OBJETIVO GENERAL.....	11
2.2. OBJETIVOS ESPECÍFICOS.....	11
2.3. RESULTADOS ESPERADOS.....	11
3. MARCO TEÓRICO Y ANTECEDENTES	12
3.1. MARCO TEÓRICO.....	12
3.1.1. Cáncer de mama	12
3.1.2. Histopatología	12
3.1.3. Imágenes histopatológicas	12
3.1.4. Segmentación de imágenes histopatológicas	12
3.1.5. Dataset CAMELYON	13
3.1.6. Métodos de normalización de color.....	13
3.1.7. Aumentado de datos.....	13
3.1.8. Software de código abierto	14
3.1.9. Modelos de <i>Deep learning</i>	14
3.2. ANTECEDENTES.....	15
4. METODOLOGÍA	18
5. ANÁLISIS DEL CONJUNTO DE IMÁGENES DE HISTOLOGÍA.....	21
6. ESTIMACIÓN DE CARGA TUMORAL MEDIANTE MODELOS DE SEGMENTACIÓN DE APRENDIZAJE MÁQUINA.....	33
7. DESEMPEÑO DE LOS MODELOS DE SEGMENTACIÓN DE IMÁGENES HISTOPATOLÓGICAS.....	51
8. CONCLUSIONES.....	58
9. TRABAJOS FUTUROS	60
10. REFERENCIAS BIBLIOGRÁFICAS	61

LISTA DE FIGURAS

Figura 4.1. Metodología propuesta.....	19
Figura 5.1. Imagen de tejido completo escaneada de muestra con máscara de verdad. (a) muestra de tejido completo con submuestreo a 256x. (b) máscara de verdad	23
Figura 5.2. Distribución de tejido frente al fondo por WSI.....	24
Figura 5.3. Porcentaje de tumor con relación a tejido por WSI	25
Figura 5.4. (a) Distribución de excentricidad y (b) Relación entre excentricidad y solidez	25
Figura 5.5. Media de color canal (a) Red, (b) Green y (c) Blue	26
Figura 5.6. Desviación estándar del color en el canal (a) Red, (b) Green y (c) Blue	27
Figura 5.7. Extracción y detección de parches de tejido en las WSI.....	29
Figura 5.8. Normalización de color del conjunto de datos. (a) Imagen de referencia, (b) parches de prueba y (c) Distribución equilibrada de colorantes.....	30
Figura 5.9. Distribución de similitud en la normalización de color.....	31
Figura 5.10. Aumentado de datos para muestras de entrenamiento. (a) Parches de tejido y (b) Máscaras de segmentación.	32
Figura 6.1. Arquitectura UNet + ResNet50 desarrollada.....	34
Figura 6.2. Evolución de la pérdida del modelo UNet + ResNet50.....	35
Figura 6.3. Evolución de la exactitud del modelo UNet + ResNet50.....	35
Figura 6.4. Evolución del AUC del modelo UNet + ResNet50.....	36
Figura 6.5. Evolución de la precisión del modelo UNet + ResNet50.....	36
Figura 6.6. Evolución del <i>recall</i> del modelo UNet + ResNet50.....	37
Figura 6.7. Evolución del <i>step</i> por época del modelo UNet + ResNet50.....	37
Figura 6.8. Arquitectura ResNet50. Tomada de [43]	39
Figura 6.9. Evolución de la pérdida del modelo ResNet50.....	39
Figura 6.10. Evolución de la exactitud del modelo ResNet50	40
Figura 6.11. Evolución del AUC del modelo ResNet50	40
Figura 6.12. Evolución de la precisión del modelo ResNet50.....	41
Figura 6.13. Evolución del <i>recall</i> del modelo ResNet50	41
Figura 6.14. Evolución del <i>step</i> por época del modelo ResNet50	42
Figura 6.15. Arquitectura U-NET. Tomada de [44].....	43
Figura 6.16. Evolución de la pérdida del modelo UNet.....	43
Figura 6.17. Evolución de la exactitud del modelo UNet	44
Figura 6.18. Evolución del AUC del modelo UNet	44
Figura 6.19. Evolución de la precisión del modelo UNet.....	45

Figura 6.20. Evolución del <i>recall</i> del modelo UNet	45
Figura 6.21. Evolución del <i>step</i> por época del modelo UNet	46
Figura 6.22. Arquitectura DenseNet121. Tomada de [45].....	47
Figura 6.23. Evolución de la pérdida del modelo DenseNet121	47
Figura 6.24. Evolución de la exactitud del modelo DenseNet121	48
Figura 6.25. Evolución del AUC del modelo DenseNet121.....	48
Figura 6.26. Evolución de la precisión del modelo DenseNet121	49
Figura 6.27. Evolución del <i>recall</i> del modelo DenseNet121	49
Figura 6.28. Evolución del <i>step</i> por época del modelo DenseNet121.....	50
Figura 7.1. Desempeño general del modelo UNet + ResNet50.....	51
Figura 7.2. Desempeño general del modelo ResNet50	52
Figura 7.3. Desempeño general del modelo UNet	52
Figura 7.4. Desempeño general del modelo DenseNet121.....	53
Figura 7.5. Análisis comparativo del rendimiento de los modelos de segmentación	54

LISTA DE TABLAS

Tabla 1. Resumen técnico del conjunto de datos CAMELYON16.....	22
Tabla 2. Transformaciones para el aumento de datos.....	31
Tabla 3. Versiones específicas de las bibliotecas principales.....	33
Tabla 4. Hiperparámetros aplicados a cada modelo	38
Tabla 5. Comparación de métricas de segmentación.....	55
Tabla 6. Estimación de la carga tumoral	56
Tabla 7. Comparación de rendimiento entre el modelo propuesto y trabajos del estado del arte publicados en revistas de alto impacto Scopus	56

INTRODUCCIÓN

En los últimos años, el cáncer de mama se ha convertido en uno de los cánceres con mayor mortalidad de mujeres de todo el mundo [1]. Es evidente el aumento acelerado de nuevos casos de cáncer de mama que han superado al cáncer de pulmón como el principal cáncer incidente en el mundo. El ganglio linfático centinela representa la primera etapa de la metástasis de los ganglios linfáticos del cáncer de mama que absorbe la linfa de un tumor primario y provoca que se diseminan las células cancerosas [2]. Por otra parte, en la práctica clínica los patólogos enfrentan desafíos en la evaluación de muestras de tejido en el microscopio debido a la heterogeneidad de las muestras que puede complicar la interpretación de las características histopatológicas, lo que induce potencialmente a la variabilidad en el diagnóstico entre patólogos [3]. Además, los avances en los tratamientos demandan análisis de tejidos más precisos y detallados.

La segmentación de imágenes histopatológicas se refiere al proceso orientado en identificar y clasificar las células o estructuras dentro de las muestras de tejido para el diagnóstico oportuno y la planificación de seguimientos. Este proceso contribuye en las actividades de los patólogos en identificar áreas de tejido canceroso y diferenciar el tipo específico de cáncer por medio de la cuantificación de características celulares con respecto al tamaño y forma de las células con el fin de reconocer la agresividad del cáncer.

Los avances y el desarrollo de técnicas de aprendizaje profundo han proporcionado variedad de métodos de análisis de imágenes histopatológicas para la identificación de metástasis de ganglios linfáticos [4]. En adición, los sistemas de diagnóstico asistido por computador desempeñan una tarea esencial en la evaluación de imágenes histopatológicas con la identificación de la región de interés que contribuye a la detección de malignidad [5].

La estimación de la carga tumoral en el cáncer de mama, mediante la segmentación automatizada de células y regiones neoplásicas en imágenes histopatológicas, ha encontrado en los modelos de aprendizaje profundo una herramienta revolucionaria. Estas arquitecturas son capaces de analizar digitaciones completas de biopsias, cuantificando con una precisión y velocidad la proporción de tejido maligno frente al estroma, el tejido adiposo y los componentes benignos circundantes. Esta cuantificación automatizada es crucial no solo para determinar el grado tumoral, sino que se erige como un biomarcador digital objetivo para pronosticar la agresividad de la enfermedad y predecir la respuesta a terapias neoadyuvantes [6]. Además, al delimitar con exactitud el tumor, estos modelos facilitan la extracción de características de la región de interés para análisis genómicos posteriores, asegurando que los resultados de pruebas como los perfiles de expresión génica sean representativos de la población celular maligna. La integración de estas herramientas en el flujo de trabajo diagnóstico no solo minimiza la variabilidad interobservador, sino que dota al patólogo de un asistente cuantitativo robusto, permitiendo una toma de decisiones más personalizada y basada en datos de cada paciente [7].

En el presente trabajo se implementó y comparó algoritmos de aprendizaje profundo enfocados en la segmentación de metástasis sobre imágenes histopatológicas de cáncer de mama para estimar carga tumoral, utilizando el conjunto de datos CAMELYON16. Para ello, se analizaron y preprocesaron los datos, incluyendo normalización de color mediante el método de Macenko y aumento de datos utilizando transformaciones geométricas y de variación de color, y se compararon cuatro arquitecturas: el modelo híbrido propuesto UNet + ResNet50, junto con ResNet50, UNet y Dense Net121. Los resultados establecieron que el modelo híbrido UNet + ResNet50 fue la solución más robusta y equilibrada, logrando

un Dice score del 94.64 % y una exactitud del 94.98 %, superando consistentemente los demás modelos, que mostraron signos de sobreajuste. Además, el modelo híbrido demostró la mayor precisión en la estimación de la carga tumoral, manteniendo un error máximo de 5.78 %, un rendimiento superior frente a los errores drásticamente altos (superiores a 20 % e incluso 35 %) cometidos por las alternativas individuales.

1. DEFINICIÓN DEL PROBLEMA

1.1. PLANTEAMIENTO DEL PROBLEMA

El cáncer de mama es considerado uno de los principales problemas en el sector salud que más afecta a mujeres con tejido mamario denso. Según cifras del Observatorio Global de Cáncer se afirma que uno de cada cuatro casos de cáncer diagnosticados en todo el mundo está relacionado con cáncer de mama y representan el quinto lugar en las causas de muerte a nivel mundial. De hecho, la incidencia del cáncer de mama fue de 23.7 por cada 100.000 habitantes durante 2018, mientras que la tasa de mortalidad reportada fue de 6.8 por cada 100.000 habitantes [8]. La metástasis del cáncer de mama representa la mayoría de las muertes por cáncer de mama, de ahí la importancia de detectar la metástasis en etapa inicial para predecir la progresión del cáncer y así obtener una comprensión profunda de la malignidad metastásica, fundamental para desarrollar intervenciones y tratamientos terapéuticos para combatir la metástasis [9].

Los patólogos analizan el tejido mamario microscópicamente y los bordes del espécimen de lumpectomía para conseguir márgenes negativos sin la presencia de cáncer. Un análisis preciso por parte del patólogo es esencial, y a su vez implica que examine entre 20 y 40 diapositivas histológicas por cada caso, resultando un proceso largo y tedioso. El análisis de imágenes histopatológicas de cáncer de mama toma gran relevancia y ayuda a los profesionales de la salud en el diagnóstico y seguimiento del proceso del tratamiento [10]. Este análisis abarca variedad de tareas y requerimientos, en donde se destaca la segmentación de imágenes histopatológicas orientada para describir detalles relacionados con la forma, volumen y ubicación de las regiones lesionadas, los cuales son esenciales para ofrecer indicadores para tomar decisiones clínicas. Sin embargo, esta tarea involucra el esfuerzo manual de expertos con alta experiencia, lo que implica gran carga de trabajo y esté sujeto a errores de interpretación a causa de las similitudes y variaciones del tumor [11].

La carga tumoral se refiere a la cantidad de tejido neoplásico presente sobre una muestra que se mide por el tamaño del tumor, número de células o extensión, y se utiliza para evaluar la severidad del cáncer y monitorear la progresión del tratamiento. La medición de la carga tumoral enfrenta varios desafíos. Los métodos tradicionales de evaluación visual de imágenes histopatológicas son subjetivos y propensos a presentar variabilidad entre observadores debido a la heterogeneidad del tejido. Además, técnicas más específicas como la tomografía por emisión de positrones o la resonancia magnética son costosas y con acceso reducido. La carga tumoral es una métrica decisiva en oncología que proporciona una visión completa del impacto físico de la enfermedad y permite una evaluación integral de la progresión y respuesta del tratamiento [12].

Por lo anterior, se han investigado métodos de segmentación automática que implementan sistemas de diagnóstico por computador que generan resultados de segmentación confiables y precisos basados en aprendizaje profundo. A su vez, estos avances en la capacidad de escaneo digital de diapositivas histopatológicas en los métodos de patología computacional podrían mejorar la eficiencia al trabajar con imágenes de diapositivas completas (WSIs) de los especímenes [13].

La segmentación de la región con metástasis en imágenes histopatológicas se considera un desafío a causa de la escasez de los conjuntos de datos anotados específicamente para este proceso y a la complejidad de

las estructuras y patrones tumorales que son evidentes. A su vez, algunos modelos computacionales pueden presentar dificultades para segmentar los variantes tipos de muestras y niveles de resolución, limitando su aplicación clínica [14]. En consecuencia, es importante desarrollar modelos avanzados para segmentar con precisión las imágenes histopatológicas y detectar la metástasis sobre las mismas.

1.2. FORMULACIÓN DEL PROBLEMA

Dicho lo anterior, surge la siguiente pregunta de investigación:

- ¿Cómo segmentar el área metastásica en imágenes histopatológicas de cáncer de mama utilizando modelos computacionales para estimar la carga tumoral?

Del mismo modo, surgen las siguientes preguntas de sistematización:

- ¿Cómo preparar el conjunto de datos de imágenes histopatológicas de cáncer de mama?
- ¿Cómo desarrollar modelos de segmentación de área metastásica en imágenes histopatológicas de cáncer de mama para estimar carga tumoral?
- ¿Cómo evaluar el desempeño de los modelos de segmentación de área metastásica en imágenes histopatológicas de cáncer de mama para estimar carga tumoral?

2. OBJETIVOS DEL PROYECTO

2.1. OBJETIVO GENERAL

Implementar modelos de aprendizaje profundo para segmentar el área metastásica en imágenes histopatológicas y estimar carga tumoral.

2.2. OBJETIVOS ESPECÍFICOS

- Analizar el conjunto de imágenes histopatológicas del cáncer de mama para obtener información relevante.
- Desarrollar modelos de segmentación de imágenes histopatológicas utilizando técnicas de aprendizaje máquina para estimar carga tumoral.
- Validar los modelos de segmentación por medio de la evaluación del desempeño utilizando métricas de segmentación.

2.3. RESULTADOS ESPERADOS

- Análisis detallado del conjunto de imágenes histopatológicas para identificar características relevantes del tejido.
- Modelos de aprendizaje computacional para la segmentación de imágenes histopatológicas.
- Estadística de validación respecto a la segmentación del área tumoral para determinar la carga tumoral.
- Documento final con la descripción de la comparación de modelos desarrollados para la segmentación de imágenes histopatológicas de cáncer de mama.

3. MARCO TEÓRICO Y ANTECEDENTES

3.1. MARCO TEÓRICO

A continuación, se presentan los conceptos principales que contextualizan la investigación desarrollada.

3.1.1. Cáncer de mama

El cáncer de mama es una enfermedad relacionada con el crecimiento sin control de células de mama. La mayoría de los tipos de cáncer de mama comienzan en los lobulillos o conductos y pueden propagarse fuera de la mama por medio de los vasos linfáticos y sanguíneos. Cuando el cáncer de mama se propaga en otras partes del cuerpo se afirma que ha hecho metástasis. Por otra parte, alrededor del 70 % y el 80 % de los cánceres de mama son relacionados con el carcinoma ductal invasivo, considerado el tipo más común de cáncer de mama, el cual provoca que las células cancerosas se propaguen en los conductos que llevan leche al pezón o por los ganglios linfáticos, incluso a otras partes del cuerpo por medio del torrente sanguíneo. Entre el 10 % y el 15 % de los cánceres de mama el carcinoma lobulillar invasivo representa el segundo tipo más común de cáncer de mama [15]. Las células cancerosas comienzan por el lobulillo y se propagan a los tejidos mamarios cercanos. También puede propagarse a otras partes del cuerpo.

3.1.2. Histopatología

La histopatología se refiere a la rama de la ciencia médica que estudia las enfermedades a partir de exámenes microscópicos de tejidos biológicos para identificar cambios o anomalías. Es fundamental en el diagnóstico de cáncer, afecciones infecciosas, autoinmunes y neurológicas, permitiendo diferenciar entre enfermedades que pueden parecer similares a simple vista y proporcionando un diagnóstico definitivo y orientado en el tratamiento. Generalmente, se obtiene una muestra del tejido durante una autopsia que se procesa y tiñe para su observación microscópica [16]. La histopatología ha evolucionado con técnicas de inmunohistoquímica, patología digital y la integración de inteligencia artificial que mejoran la precisión y eficiencia diagnóstica [17].

3.1.3. Imágenes histopatológicas

Las imágenes histopatológicas se refieren a imágenes de muestras microscópicas de tejido que son utilizadas para analizar y diagnosticar enfermedades relacionadas con el cáncer. Los patólogos examinan la estructura del tejido, la forma y la distribución celular para determinar si son malignas o benignas. Además, los patólogos utilizan colorantes, en donde la hematoxilina y la eosina son los más utilizados para clasificar las estructuras histopatológicas celulares en muestras de tejido [18]. La hematoxilina se aprecia de color azul o púrpura, y la eosina aparece de color rosado. En la segmentación de imágenes histopatológicas, se suelen utilizar modelos de redes profundas y arquitectura de redes neuronales convolucionales para analizar imágenes de histopatología y extraer información de las características del tejido celular.

3.1.4. Segmentación de imágenes histopatológicas

La segmentación de imágenes consiste en dividir una imagen de un tejido biológico en secciones relevantes para facilitar su análisis e interpretación. Este proceso permite reconocer y clasificar estructuras celulares y tisulares en zonas del tejido. Este proceso es crucial para el análisis y diagnóstico de patología ya que permite identificar características morfológicas de las células que facilita la detección del cáncer [19]. La

segmentación de imágenes histopatológicas puede realizarse mediante técnicas y métodos tradicionales basados en umbrales y con modelos de aprendizaje automático y redes neuronales convolucionales que permiten una segmentación automatizada robusta y precisa.

3.1.5. Dataset CAMELYON

En el contexto de segmentación médica de imágenes H&E (hematoxilina y eosina), un dataset se refiere a una colección de imágenes digitales de tejidos teñidos, junto con anotaciones manuales o automáticas que marcan regiones de interés, como áreas patológicas o núcleos celulares. Los datasets son el componente principal para entrenar, validar y comparar algoritmos de segmentación, principalmente los basados en aprendizaje profundo, que requieren grandes volúmenes de datos etiquetados para aprender a identificar y separar estructuras relevantes en las imágenes [20]. El dataset CAMELYON es uno de los recursos más ampliamente utilizados en la patología computacional para el análisis de metástasis en ganglios linfáticos de pacientes con cáncer de mama. Se incluyen 1,399 imágenes de portaobjetos completos (WSIs) teñidos con H&E que fueron recolectadas de cinco centros médicos diferentes con una amplia variedad en la apariencia y tinción de las muestras. El conjunto completo ocupa aproximadamente 3 terabytes de datos y 209 WSIs cuentan con contornos manuales detallados de todas las metástasis [21]. Este dataset ha sido fundamental para el desarrollo y evaluación de algoritmos de inteligencia artificial para automatizar la detección de células tumorales en imágenes histopatológicas digitales.

3.1.6. Métodos de normalización de color

La normalización de color en H&E busca reducir la variabilidad interlaboratorio estandarizando la apariencia cromática respecto a una imagen de referencia, mejorando así la robustez de algoritmos de segmentación. Los métodos más utilizados se dividen en enfoques globales estadísticos y basados en separación de tinciones. Dentro de los enfoques de normalización de color, el método de Macenko se ha consolidado como la referencia más utilizada entre los métodos basados en separación de tinciones, que trabaja en el espacio de densidad óptica (OD), estima automáticamente los vectores de color de hematoxilina y eosina mediante descomposición en valores singulares (SVD) y, posteriormente, reescala las intensidades de tinción de la imagen fuente para que coincidan con las de una imagen de referencia, logrando una representación cromática más homogénea entre preparaciones H&E de distintos laboratorios. En contraste, Reinhard transforma la imagen al espacio $\alpha\beta$ y ajusta globalmente media y desviación estándar de cada canal para igualarlas a las de referencia, lo que suele producir colores naturales y buena conservación morfológica con baja complejidad computacional. Por su parte, Vahadane mantiene el esquema de trabajo en OD y la separación H&E de Macenko, pero reemplaza la SVD por una factorización matricial no negativa dispersa (sparse NMF) para obtener una descomposición más robusta de las componentes de tinción y preservar la estructura tisular [22].

En [23] se muestra que, si bien Reinhard suele ser competitivo y estable entre tejidos, y Vahadane puede ofrecer mejor modelado de mezclas de tinción, ambos pueden introducir artefactos cromáticos, y su rendimiento relativo frente a Macenko depende del tejido, del conjunto de datos y de la tarea (clasificación, segmentación, detección), por lo que en la práctica Macenko suele usarse como punto de partida y *baseline* frente al resto de métodos.

3.1.7. Aumentado de datos

El aumento de datos es una técnica fundamental en el aprendizaje profundo que permite incrementar la cantidad, calidad y diversidad del conjunto de entrenamiento mediante transformaciones o generación

artificial de nuevas muestras a partir de los datos originales. Esta técnica ayuda a mitigar el sobreajuste al exponer al modelo a una mayor variedad de variaciones y mejorar su capacidad de generalización [24]. Las transformaciones incluyen rotaciones, desplazamientos, cambios de escala, ajustes de brillo o contraste en imágenes, así como la adición de ruido controlado. Los modelos computacionales se vuelven más robustos ante variaciones del entorno real al diversificar los datos, indicando un mejor rendimiento en tareas de clasificación, segmentación o detección de objetos [25].

3.1.8. Software de código abierto

El software de código abierto se caracteriza por tener su código fuente accesible al público, permitiendo su libre modificación y distribución. Este modelo promueve la colaboración y transparencia entre desarrolladores, facilitando el intercambio de soluciones tecnológicas y la mejora continua. En el ámbito médico, el código abierto ha adquirido relevancia por su accesibilidad y adaptabilidad, siendo utilizado en sistemas de gestión de información clínica, plataformas de interoperabilidad sanitaria y herramientas de análisis de datos médicos [26]. Gracias a esta flexibilidad, los profesionales de la salud pueden personalizar las aplicaciones según sus requerimientos, optimizando la atención del paciente y fomentando la innovación colaborativa.

3.1.9. Modelos de *Deep learning*

Los modelos de Deep learning han evolucionado la segmentación de imágenes histológicas teñidas, permitiendo identificar y delimitar estructuras celulares y tisulares con alta eficiencia. Las redes neuronales convolucionales (CNN) son la base de la mayoría de los modelos de segmentación y permiten el aprendizaje automático de características relevantes y la clasificación de regiones, superando métodos tradicionales en robustez [27]. ResNet50 es una red neuronal convolucional profunda de 50 capas que utiliza bloques residuales para simplificar el entrenamiento de arquitecturas muy profundas. Estos bloques introducen conexiones de salto (skip connections) para que el gradiente fluya directamente a través de la red, mitigando el problema del gradiente evanescente. La arquitectura se compone de cinco etapas: una capa inicial de convolución y max-pooling, seguida de cuatro etapas con bloques residuales (3, 4, 6 y 3 bloques respectivamente), donde cada bloque contiene tres capas convolucionales (1x1, 3x3 y 1x1) con normalización por lotes (BatchNorm) y activación ReLU. Por último, se incluye una capa average pooling y una fully connected para clasificación [28]. UNet es una arquitectura de red neuronal convolucional ampliamente utilizada en problemas de segmentación semántica de imágenes biomédicas, en donde se asigna una etiqueta a cada píxel para un análisis detallado de la composición de la imagen. UNet se compone de codificador y un decodificador, en donde el codificador son una serie de capas convolucionales que reducen el tamaño de las imágenes, pero incrementa el número de canales con el fin de extraer el mayor número de características, identificando la información más relevante. En el decodificador se realiza una serie de capas de deconvolución para producir una salida con las mismas dimensiones de la entrada a medida que se ejecutan las capas de deconvolución con una fusión de la imagen con las características extraídas por el codificador [29]. Por otra parte, DenseNet121 es una arquitectura de red neuronal convolucional que se caracteriza por su diseño de conexiones densas entre capas, donde cada capa está conectada a todas las capas siguientes en un bloque denso, promoviendo la reutilización de características y mejorando el flujo de gradientes durante el entrenamiento. El modelo consta de 121 capas que están organizadas en cuatro bloques densos intercalados con capas de transición que reducen las dimensiones espaciales utilizando el pooling para controlar la complejidad computacional. Cada capa dentro de un

bloque denso aplica operaciones convolucionales (Conv2D + BatchNorm + ReLU) y concatena sus salidas con las capas anteriores, incrementando progresivamente el número de canales [30].

3.2. ANTECEDENTES

Asimismo, se presentan investigaciones enfocadas en la segmentación de la región de interés en imágenes histopatológicas de cáncer de mama.

- **Breast Cancer Histopathological Images Segmentation Using Deep Learning [31]**

Este antecedente presenta dos enfoques de segmentación semántica en imágenes histopatológicas de cáncer de mama. En primer lugar, se propone una arquitectura auto codificada para segmentación no supervisada y, por otro lado, propone una arquitectura U-NET para segmentación supervisada. Los resultados muestran que el modelo U-NET entrenado con anotaciones manuales y segmentación no supervisada proporcionó un rendimiento con valores de intersección sobre unión de 86,1 % y 84,8 % que permiten concluir que el método propuesto no supervisado muestra un rendimiento de segmentación celular comparable en comparación con el modelo U-NET modificado. Estos enfoques ofrecen alternativas robustas para la segmentación automática en contextos con o sin anotaciones, facilitando el análisis de imágenes H&E de cáncer de mama.

- **Multi-module UNet++ for colon cancer histopathological image segmentation [32]**

El antecedente presenta RPAU-Net++, un modelo avanzado basado en UNet++ que integra un codificador ResNet50, un módulo de fusión piramidal conjunta (JPFM) con convoluciones dilatadas y un módulo de atención dual (CBAM) para mejorar la segmentación de imágenes histopatológicas de cáncer colorrectal. Este enfoque aborda desafíos relacionados con la variabilidad morfológica, la indeterminación en los bordes y el ruido de fondo, logrando una mejor extracción y fusión multiescala de características junto con atención espacial y de canal para enfocar regiones relevantes. Evaluado en los conjuntos de datos GlaS, CoNIC y PanNuke, RPAU-Net++ supera a los modelos UNet, UNet++ y nnUNet, alcanzando mejoras en las métricas IoU hasta un 4.23 % y Dice score de 3.50 % sobre U-Net en GlaS, además de mostrar eficiencia computacional mejorada y estabilidad, lo que la posiciona como una solución eficiente para aplicaciones clínicas en segmentación histopatológica.

- **BreasTDLUSeg: A coarse-to-fine framework for segmentation of breast terminal duct lobular units on histopathological whole-slide images [33]**

En este antecedente se introduce a BreasTDLUSeg, un marco de dos etapas que utiliza una red de clasificación de niveles y una red de segmentación de píxeles para la segmentación automática de las unidades lobuladas de los conductos terminales (TDLU) en imágenes histopatológicas basándose en la atención multi-escala que contribuye en el análisis a nivel patológico del cáncer de mama. Esta propuesta superó a otros métodos con coeficientes de similitud de Dice de 79.97 % y 92.93 % sobre dos conjuntos de datos de cáncer de mama construidos que contienen 530 imágenes con anotaciones a nivel de parche y 2,322 imágenes con anotaciones de nivel de píxeles. Se aborda efectivamente los desafíos clave de la segmentación de TDLUs, como la heterogeneidad, el desequilibrio de clases y la eficacia computacional.

- **H2G-Net: A multi-resolution refinement approach for segmentation of breast cancer region in gigapixel histopathological images [34]**

En el documento se expone el diseño de una red neuronal convolucional en cascada H2G-Net para segmentar el cáncer de mama en imágenes histopatológicas gigapixel involucrando una etapa de detección utilizando método de parches y una etapa de refinamiento mediante un autoencoder convolucional para mejorar la precisión. Se evaluó la segmentación de tumores en 624 imágenes de diapositivas de cáncer de mama. Se logró un coeficiente de similitud de Dice de 0.933 ± 0.069 en un conjunto de imágenes de prueba de 90 imágenes, superando a los enfoques con única resolución. El modelo se integra en la plataforma FastPathology, permitiendo su uso en entornos clínicos. Estas innovaciones mejoran en precisión, escalabilidad y accesibilidad del diagnóstico digital en patología mamaria.

- **Out-of-distribution generalization for segmentation of lymph node metastasis in breast cancer** [35]

Se presenta un enfoque para mejorar la generalización de modelos de segmentación de metástasis en ganglios linfáticos en cáncer de mama, abordando el problema de cambio de dominio (OOD) que afecta el rendimiento de los modelos entrenados en datos específicos. Se utilizan técnicas no supervisadas de agrupamiento y muestreo para crear conjuntos de datos de entrenamiento representativos a nivel de parches, basándose en extractores de características preentrenados con imágenes histopatológicas e ImageNet. Para evaluar la capacidad de generalización entre datos conocidos y desconocidos se propone el método estadístico TOST para medir la equivalencia entre dominios usando márgenes dependientes de los datos. Los modelos que demostraron generalización se combinaron en ensambles que alcanzaron una puntuación F1 de 0.81, destacando el potencial clínico de este enfoque. Además, se establece un precedente para el uso de TOST en tareas de generalización en visión computacional médica, proponiendo nuevas estrategias de curación de datos basadas en características histológicas reales, no solo en heurísticas visuales.

- **Region Segmentation of Whole-Slide Images for Analyzing Histological Differentiation of Prostate Adenocarcinoma Using Ensemble EfficientNetB2 U-Net with Transfer Learning Mechanism** [36]

En este antecedente, se implementó un modelo de Deep learning con aprendizaje por transferencia para segmentar áreas cancerosas en imágenes de muestra de tejido (WSI). Se utilizó un modelo Ensemble U-Net para identificar y delimitar regiones de estroma, tejido benigno y tejido canceroso. El conjunto de datos de acceso público se obtuvo del repositorio Kaggle, incluyendo 1000 imágenes WSI, de las cuales se extrajeron 8100 parches para entrenamiento y 900 para prueba. El modelo alcanzó un coeficiente de Dice de 0.891, una intersección sobre la unión (IoU) de 0.811 y una distancia de Hausdorff de 15.9 en el conjunto de prueba. Esta herramienta mejora la capacidad del patólogo para predecir el desarrollo de la enfermedad y optimizar la eficacia del tratamiento al aislar las regiones cancerosas en las imágenes.

- **RGSB-UNet: Hybrid Deep Learning Framework for Tumour Segmentation in Digital Pathology Images** [37]

En este antecedente, se introduce RGSB-UNet, un marco de aprendizaje profundo híbrido enfocado para la segmentación de tumores en imágenes WSI. Combina bloques residuales Ghost con normalización conmutable (RGS) y Bottleneck Transformer (BoT) dentro de una arquitectura UNet, lo que permite capturar simultáneamente correlaciones espaciales locales y dependencias globales de características. Utilizando del dataset GlAS, el modelo logra un Dice de 0.8865, una precisión de píxeles de 0.9823 y un índice de

Jaccard de 0.7953. Las contribuciones del modelo incluyen una mejor estabilidad, lograda mediante el uso de normalización conmutable, así como una capacidad de refinamiento superior en la delimitación de los bordes durante la segmentación.

4. METODOLOGÍA

La presente investigación se enmarca en un diseño aplicado [38] y experimental comparativa [39], orientado al desarrollo y validación de modelos de segmentación de imágenes histopatológicas mediante técnicas de aprendizaje profundo, con el fin de estimar la carga tumoral en cáncer de mama. Se considera investigación aplicada por su enfoque en la resolución de una necesidad concreta en el ámbito clínico, y experimental comparativa porque se evaluarán distintos modelos bajo condiciones controladas.

El paradigma adoptado es el positivista cuantitativo, ya que parte de la premisa de que los fenómenos asociados a la morfología tumoral en imágenes pueden medirse objetivamente y analizarse mediante técnicas de desempeño especializadas. El enfoque es cuantitativo, al basarse en el procesamiento computacional de datos y la obtención de resultados expresados en valores numéricos.

El diseño de investigación corresponde a un estudio experimental, transversal y comparativo. Es experimental porque se manipulan variables independientes (tipo de modelo de segmentación y hiperparámetros), transversal porque se evalúa el desempeño de los modelos en un momento específico, y comparativo porque se contrastan los resultados de distintas arquitecturas de aprendizaje profundo.

Para la recolección y preparación de datos, se utilizó el conjunto de imágenes de histología pública CAMELYON16, especializado en metástasis de cáncer de mama en ganglios linfáticos. La ventaja de este recurso es que incluye anotaciones de referencia o *ground truth* realizadas por patólogos expertos. Estas delimitaciones de las áreas metastásicas, almacenadas en archivo XML, constituyeron la base de verdad para el proceso de aprendizaje profundo posterior.

El procedimiento metodológico se desarrolló en etapas sucesivas. Inicialmente, se preprocesaron estas anotaciones para convertirlas en etiquetas utilizables por los modelos. Utilizando el software ASAP (Automated Slide Analysis Platform), se cargaron las imágenes de tejido completo (WSI) junto con sus archivos XML de anotación. El software permitió visualizar las anotaciones superpuestas al tejido y generar a partir de ellas máscaras de segmentación binarias en formato TIFF, donde las áreas tumorales quedaron codificadas en un canal y el tejido sano o fondo en otro.

A continuación, se realizó el preprocesamiento tanto de las imágenes originales como de las máscaras generadas. Dada la enorme resolución de las WSI y sus correspondientes máscaras, se aplicó una estrategia de división de parches de tamaño 256x256 píxeles, asegurando la correspondencia píxel a píxel entre la imagen de tejido y su etiqueta. A los parches de imagen se les aplicó normalización de color utilizando el método de Macenko para reducir la variabilidad debida a tinciones diferenciales y técnicas de aumento de datos para mejorar la robustez de los modelos.

Posteriormente, se procedió al entrenamiento y validación de varios modelos de segmentación semántica, incluyendo arquitecturas como UNet con encoder ResNet50, ResNet50, UNet y DenseNet121. Todo el procesamiento computacional y el entrenamiento se ejecutaron en el entorno especializado Google Colab, aprovechando la máquina NVIDIA Tesla T4 GPU (Unidad de Procesamiento Gráfico). Finalmente, el desempeño de los modelos se validó en un conjunto de prueba independiente, y la segmentación predicha por cada modelo se utilizó para estimar la carga tumoral, comparando estos resultados cuantitativamente con las máscaras de referencia generadas por ASAP a partir de las anotaciones originales de los patólogos.

El análisis de datos se realizó mediante la evaluación de métricas de segmentación (exactitud, sensibilidad, precisión, F1 score, Dice score, Jaccard y AUC) y el cálculo del error absoluto para determinar diferencias sobre la estimación porcentual de la carga tumoral entre los modelos comparados.

La metodología propuesta para la comparación de modelos segmentación de imágenes histopatológicas de cáncer de mama para estimar carga tumoral se resume en la figura 4.1. Se presentan 3 etapas: el análisis de imágenes histopatológicas del conjunto de datos CAMELYON16, la estimación de la carga tumoral mediante modelos de segmentación de aprendizaje máquina y el desempeño de los modelos de segmentación de imágenes histopatológicas.

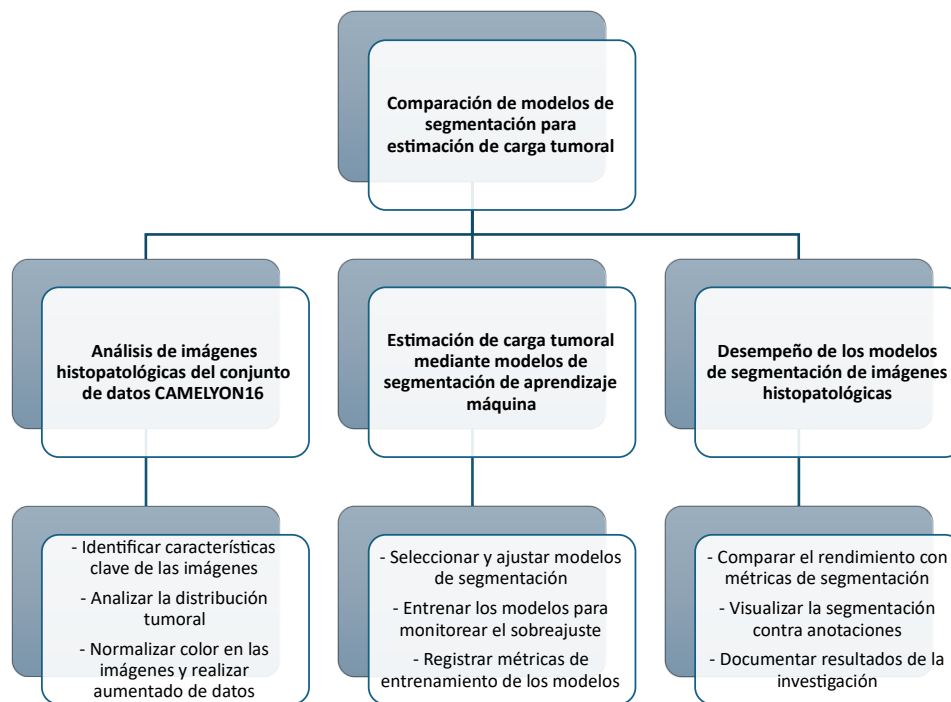


Figura 4.1. Metodología propuesta

En la primera etapa de la metodología, se realizó un análisis de las imágenes histopatológicas del conjunto de datos CAMELYON16, identificando las características clave, llevando a cabo un análisis exploratorio de datos para comprender la distribución y propiedades de las regiones tumorales, y aplicando técnicas de preprocesamiento para la extracción de parches de 256 x 256 píxeles, normalización de color y aumento de datos para mejorar la generalización del modelo. La segunda etapa se centró en estimar la carga tumoral mediante modelos de segmentación basados en aprendizaje profundo que se diseñaron para delimitar las áreas afectadas. Finalmente, en la última etapa, se evaluó el desempeño de los modelos comparando su capacidad de segmentación y la exactitud en la estimación de la carga tumoral a través de métricas específicas, con la documentación detallada de los resultados obtenidos durante la investigación.

- Objetivo 1: Analizar el conjunto de imágenes histopatológicas del cáncer de mama para obtener información relevante.

Actividades:

1. Identificar las características clave de las imágenes histopatológicas del conjunto de datos.
 2. Realizar un análisis exploratorio de datos para comprender la distribución y características de las regiones tumorales en las imágenes histopatológicas.
 3. Realizar un preprocesamiento de normalización de color y utilizar un método de aumento de datos para mejorar la generalización.
- Objetivo 2: Desarrollar modelos de segmentación de imágenes histopatológicas utilizando técnicas de aprendizaje máquina para estimar carga tumoral.

Actividades:

1. Seleccionar los modelos de segmentación y ajustar hiperparámetros para implementación.
 2. Entrenar modelos de segmentación seleccionados y monitorear el desempeño para evitar sobreajuste.
 3. Registrar métricas e información relevante de la evolución del proceso de entrenamiento de los modelos de segmentación seleccionados.
- Objetivo 3: Validar los modelos de segmentación por medio de la evaluación del desempeño utilizando métricas de segmentación.

Actividades:

1. Comparar el rendimiento de la segmentación y la precisión de la estimación de la carga tumoral a cada modelo seleccionado utilizando métricas cuantitativas.
2. Visualizar los resultados de la segmentación junto con las anotaciones verdaderas.
3. Documentar los resultados de la investigación.

5. ANÁLISIS DEL CONJUNTO DE IMÁGENES DE HISTOLOGÍA

La selección del dataset CAMELYON16 para la segmentación de cáncer de mama en imágenes H&E se justifica por ser un conjunto de datos público reconocido y ampliamente utilizado que garantiza la comparabilidad con el estado del arte [40]. Contiene imágenes de portaobjetos completos de alta resolución de dos centros médicos, lo que proporciona una heterogeneidad esencial en la preparación tisular y el teñido para entrenar modelos robustos y generalizables.

Además, sus anotaciones detalladas de regiones metastásicas realizadas por expertos permiten un entrenamiento y evaluación precisas de modelos de segmentación supervisada, mientras que su estructura multiresolución y la presencia de metástasis de diversos tamaños lo convierten en un recurso para abordar los desafíos de escala y detección de lesiones pequeñas inherentes a esta tarea [41].

Se aplica un análisis a las imágenes de lámina completa (WSI) del conjunto de datos CAMELYON16 para segmentar regiones en ganglios linfáticos y estimar la carga tumoral. La caracterización del conjunto de datos se inició con la identificación de sus anotaciones patológicas disponibles, las cuales constituyen el estándar de referencia para el entrenamiento y validación de los modelos. Las anotaciones fueron proporcionadas por el conjunto de datos en archivos XML que contienen los vértices de los polígonos que delimitan las regiones de interés en las imágenes histopatológicas, obteniendo a partir de ellas máscaras binarias a nivel de píxel.

Por otro lado, se analizaron propiedades clave de las imágenes de lámina completa (WSI) para cuantificar su variabilidad inherente. Se realizó una segmentación inicial para diferenciar la distribución del tejido frente al fondo (áreas sin tejido). A partir de las anotaciones en XML, se calculó el porcentaje de tumor con relación al tejido total segmentado, una métrica directa de la carga tumoral. Además, se evaluaron características morfológicas de los focos tumorales mediante el análisis de la distribución de excentricidad (grado de elongación de una región) y la relación entre excentricidad y solidez (medida de compacidad y convexidad).

Para caracterizar las propiedades de color y tinción, se calculó la media del canal Rojo, Verde y Azul (RGB) para cada WSI, así como la distribución de la desviación estándar de estos canales, parámetros que reflejan la tendencia central del color y su variabilidad o contraste interno.

Por otra parte, dado al tamaño de las WSI se implementa una estrategia de extracción de parches para procesar regiones de interés y asegurar una representación adecuada tanto del tejido sano como tumoral. Posteriormente, se definen criterios de balanceo para asegurar una distribución apropiada del tejido normal y tumoral. Seguidamente, se aplica normalización de tinción a los parches extraídos. Además, se utiliza aumento de datos incorporando transformaciones para mejorar la diversidad del conjunto de entrenamiento. Finalmente, se consolidan los datos para el desarrollo de los modelos computacionales.

El conjunto de datos CAMELYON16 se refiere a una colección de libre acceso de imágenes histopatológicas digitalizadas de ganglios linfáticos teñidos con hematoxilina y eosina (H&E), recolectadas para el estudio de metástasis en pacientes con cáncer de mama. Estas imágenes se digitalizan mediante escáneres de alta resolución para permitir un análisis profundo a nivel celular. El conjunto de datos incluye 400 imágenes de diapositivas completas almacenadas en formato TIFF (Formato de Archivo de Imagen Etiquetada), con profundidad de color en canales rojo, verde y azul [42].

El conjunto de datos evidencia variaciones en la tinción de las imágenes, causadas por diferencias en los procesos de tinción H&E, la calidad de los reactivos y los parámetros de digitalización. Estas diferencias representan un desafío para el análisis automatizado, afectando el rendimiento de los modelos de segmentación [43]. La Tabla 1 resume el conjunto de datos incluyendo aspectos técnicos y descriptivos de las imágenes de histopatología.

Tabla 1. Resumen técnico del conjunto de datos CAMELYON16

Aspecto	Detalle
Tipo de datos	Imágenes histopatológicas digitalizadas
Fuente	Ganglios linfáticos de pacientes con cáncer de mama
Resolución	0.24 μm /pixel (escaneo a 40x)
Formato de imagen	.tiff (WSIs) y .xml (anotaciones)
Tamaño	1,160 GB
Dimensión de imagen	Aproximadamente 100,000 x 100,000 píxeles
Profundidad de color	24 bits (RGB)
Número de láminas	400
Técnica de tinción	Hematoxilina y eosina (H&E)

Para la generación de las máscaras reales de segmentación se emplearon las anotaciones en formato en XML proporcionadas por el conjunto de datos y que delimitan las regiones de interés en las imágenes histopatológicas. Estas anotaciones fueron procesadas y visualizadas utilizando el software especializado ASAP y se llevó a cabo la conversión de las anotaciones vectoriales en mapas binarios con resolución a nivel de píxel que permitió la representación de las áreas tumorales y no tumorales.

En la figura 5.1, la sección (a) presenta una imagen de tejido completo escaneada la cual representa una muestra histológica de alta resolución utilizada en el diagnóstico médico con submuestreo a 256x. Esta imagen captura detalles microscópicos relacionados con morfología celular y organización tisular, que son esenciales para la identificación de patologías. Por otra parte, la sección (b) muestra la máscara de verdad correspondiente, generada en el software especializado con anotaciones expertas que delimitan regiones tumorales para entrenar modelos computacionales en tareas de segmentación.

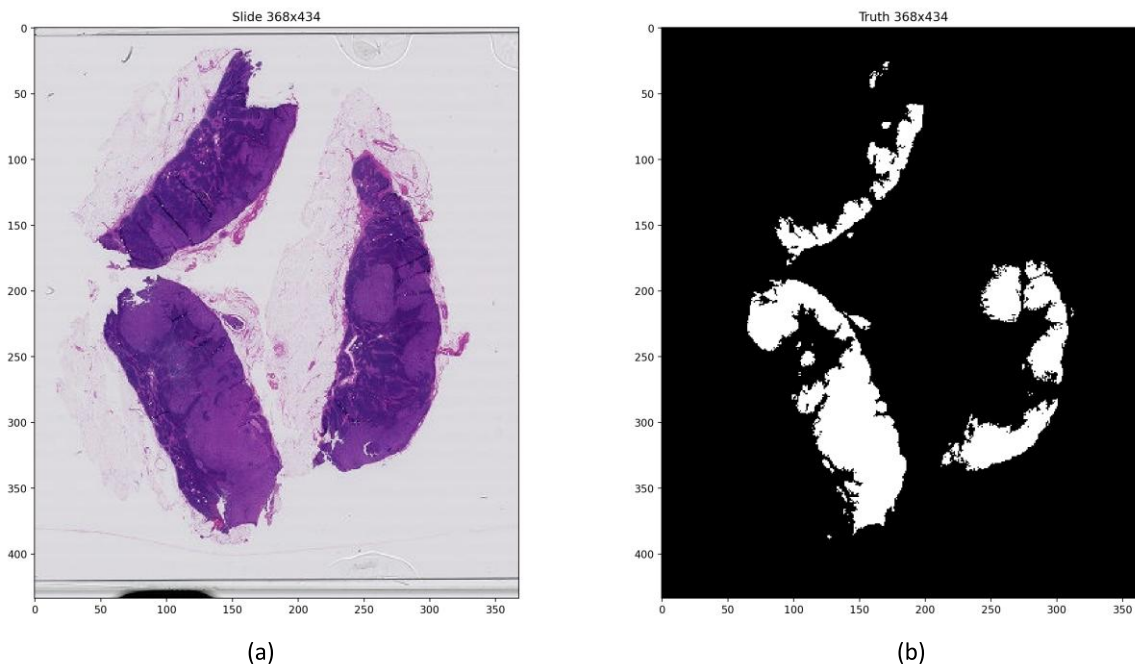


Figura 5.1. Imagen de tejido completo escaneada de muestra con máscara de verdad. (a) muestra de tejido completo con submuestreo a 256x. (b) máscara de verdad

Se aplica un método automatizado para la identificación y clasificación de parches de tejido en 10 imágenes de portaobjetos completo en patología, discriminando entre regiones tumorales y normales mediante un esquema de filtrado multietapa. Esta fase permite eliminar progresivamente áreas irrelevantes, artefactos y tejido no informativo, y enfocarse en las zonas diagnósticamente significativas. La selección de 10 WSI se justifica por su capacidad de proporcionar una evaluación robusta, manteniendo un balance entre diversidad de patrones tisulares y la viabilidad del procesamiento computacional, cubriendo variabilidad biológica sin comprometer el tiempo de análisis, siendo esto crucial en etapas tempranas del desarrollo de herramientas de patología digital.

La distribución de tejido frente al fondo en las WSI se presenta a continuación en la figura 5.2, en donde se muestra una gran variabilidad en la proporción de área ocupada por tejido frente al fondo. En efecto, el área del tejido es significativamente menor que el área de fondo en todas las muestras, con valores que oscilan entre 15,185 y 52,376, a diferencia que, el fondo abarca desde 118,277 hasta 308,028.

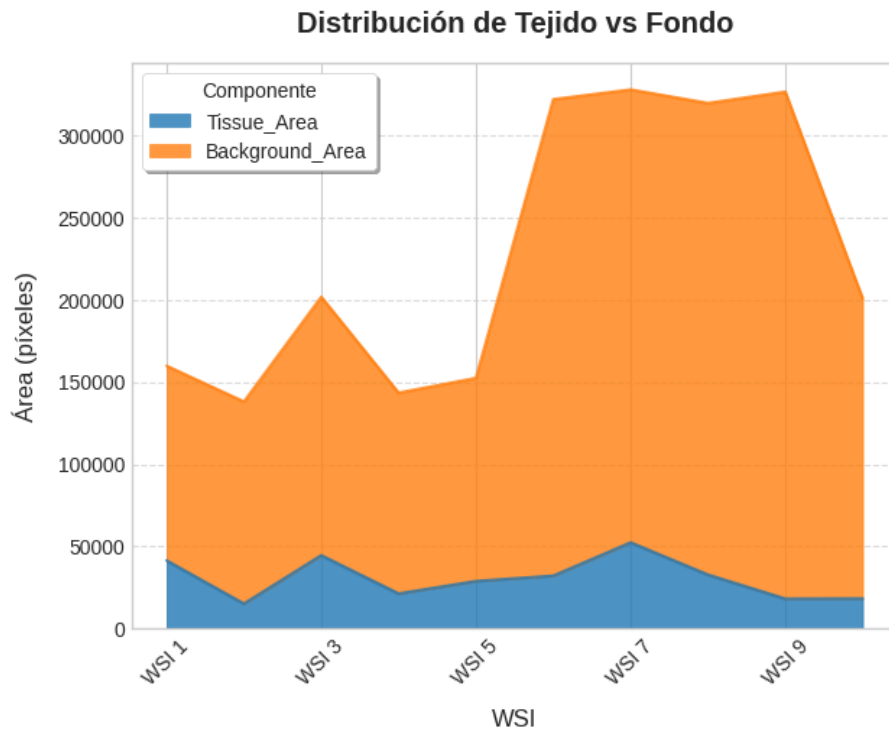


Figura 5.2. Distribución de tejido frente al fondo por WSI

El análisis de muestras de tejido en WSI releva una variabilidad en el porcentaje de área tumoral con relación al área total de tejido. Como se presenta en la figura 5.3, los valores oscilan entre externos, desde muestras con infiltración tumoral mínima (3.9 % y 4.1 %) hasta casos con predominio absoluto de tumor (98.4 % y 97.8 %). Se observa una distribución heterogénea donde algunas muestras presentan tejido tumoral moderado (17.5 %, 20.1 %, 37.7 %, 37.60 %, 41.8 % y 42.0 %), sin concentrarse en un rango específico. Esta dispersión indica que las muestras proceden de diferentes contextos patológicos, reflejando tanto tumores en etapas tempranas con baja carga tumoral como casos avanzados con reemplazos masivos del tejido normal, lo que resulta en una variación en las características entre las diferentes muestras, lo que se conoce como alta heterogeneidad intermuestral.

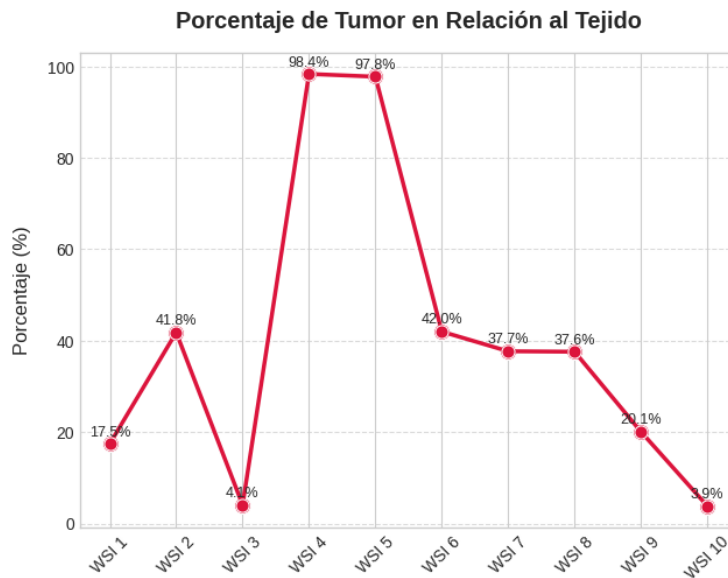


Figura 5.3. Porcentaje de tumor con relación a tejido por WSI

El análisis de las regiones en las 10 imágenes WSI revela una distribución bimodal de excentricidad como se presenta en la figura 5.4 (a), con una concentración masiva de formas perfectamente circulares (excentricidad = 0 en 282 regiones) y formas perfectamente lineales (excentricidad = 1 en 1,116 regiones), mientras que solo 201 regiones muestran valores intermedios. Esta división se correlaciona directamente con la solidez, donde la gran mayoría de las regiones (1,401) presentan una solidez máxima de 1.0, indicando formas compactas y bien definidas sin concavidades. La relación mostrada en la figura 5.4 (b) muestra que tanto las regiones circulares como las lineales comparten una alta solidez, sugiriendo que las estructuras tisulares analizadas, independientemente de su forma elongada o redondeada, tienen a presentar contornos convexos y bien delimitados, lo que podría corresponder morfológicamente a patrones histológicos definidos como nidos tumorales compactos o estructuras glandulares preservadas.

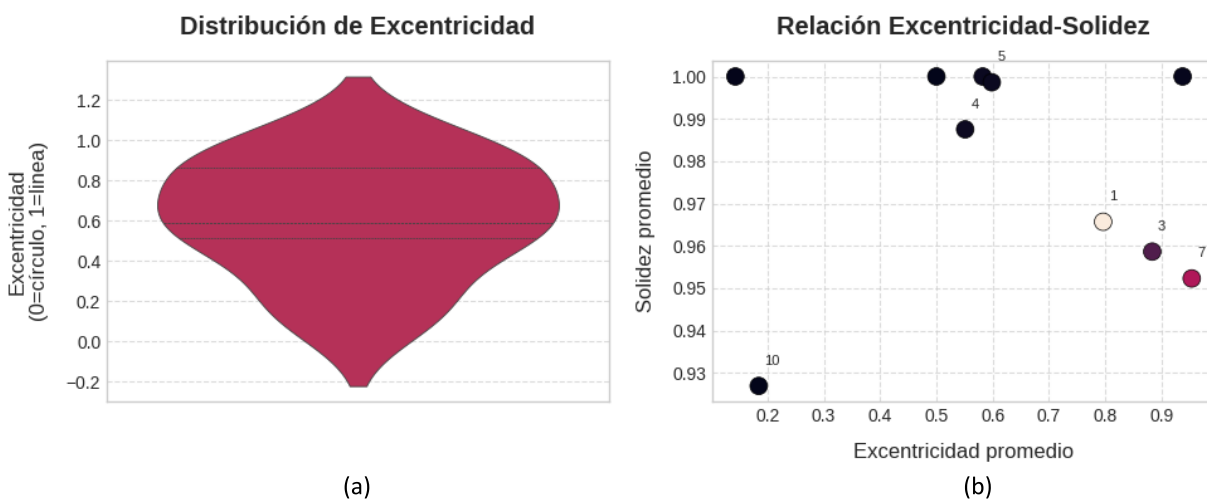


Figura 5.4. (a) Distribución de excentricidad y (b) Relación entre excentricidad y solidez

El análisis de las 10 imágenes WSI muestra valores medios específicos en regiones tumorales. Como se presenta en la figura 5.5, la WSI 9 presenta la intensidad más alta en todos los canales (Red: 223.87, Green: 214.45 y Blue: 220.87), mientras WSI 5 muestra los valores más bajos (Red: 192.99, Green: 177.24 y Blue: 203.27). En regiones no tumorales, WSI 9 también registra los valores más altos (Red: 221.13, Green: 214.54 y Blue: 218.95) y WSI 1 los más bajos (Red: 185.35, Green: 165.35 y Blue: 196.15).

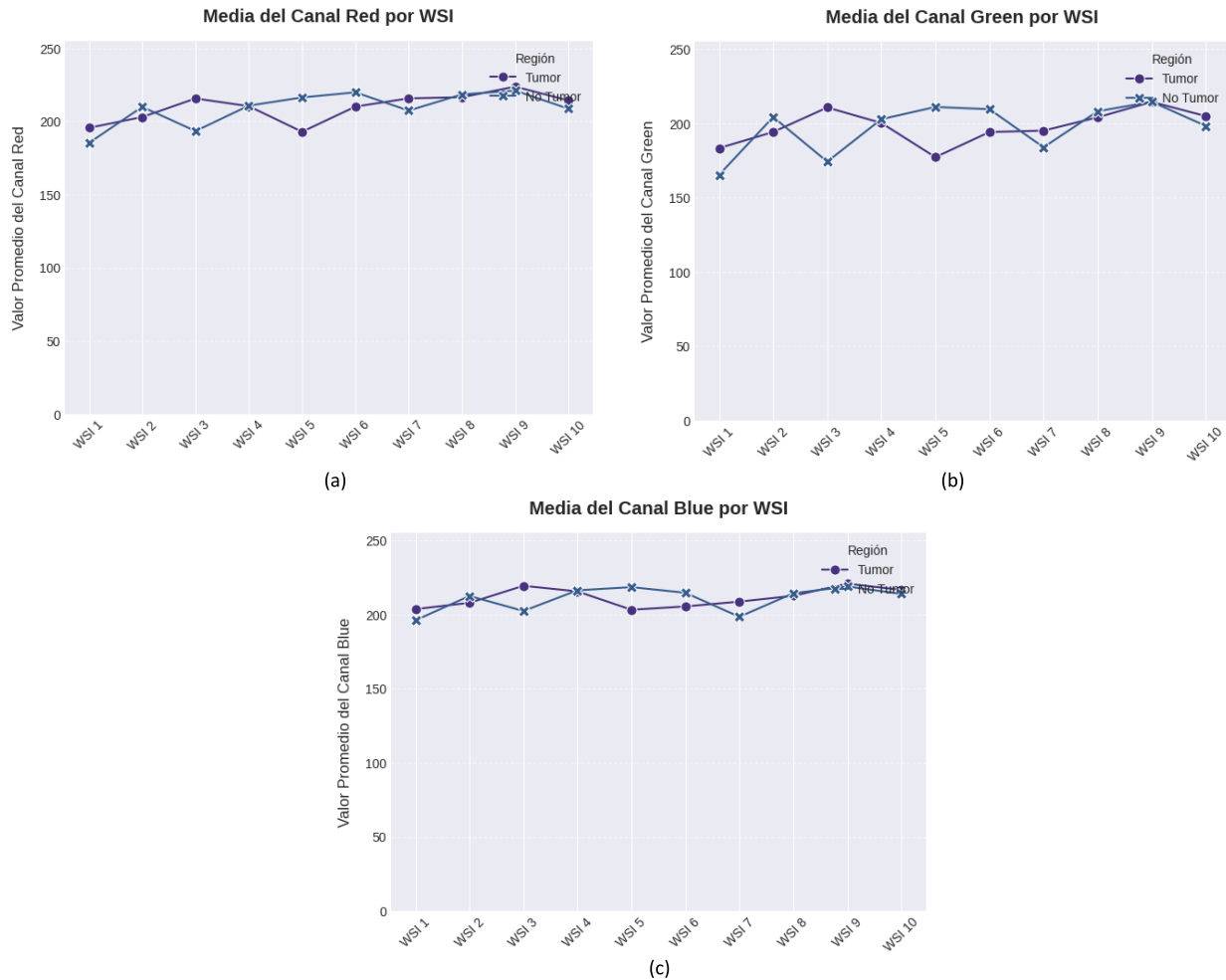


Figura 5.5. Media de color canal (a) Red, (b) Green y (c) Blue

Respecto a la distribución de variabilidad (figura 5.6), el tejido tumoral exhibe las mayores desviaciones estándar en WSI 6 (Red: 51.99, Green: 67.00 y Blue: 53.49), mientras que el tejido no tumoral muestra extrema homogeneidad en WSI 5 (Red: 12.99, Green: 18.26 y Blue: 11.41) y máxima heterogeneidad en WSI 1 (Red: 54.23, Green: 75.38 y Blue: 41.25). Individualmente, el canal Green demuestra la mayor variabilidad general en ambos tipos de tejido, con desviaciones que alcanzan 75.38 en no tumor (WSI 1) y 67.00 en tumor (WSI 6), lo cual confirma que el tejido tumoral presenta consistentemente mayor heterogeneidad en la distribución de color que el tejido no tumoral.

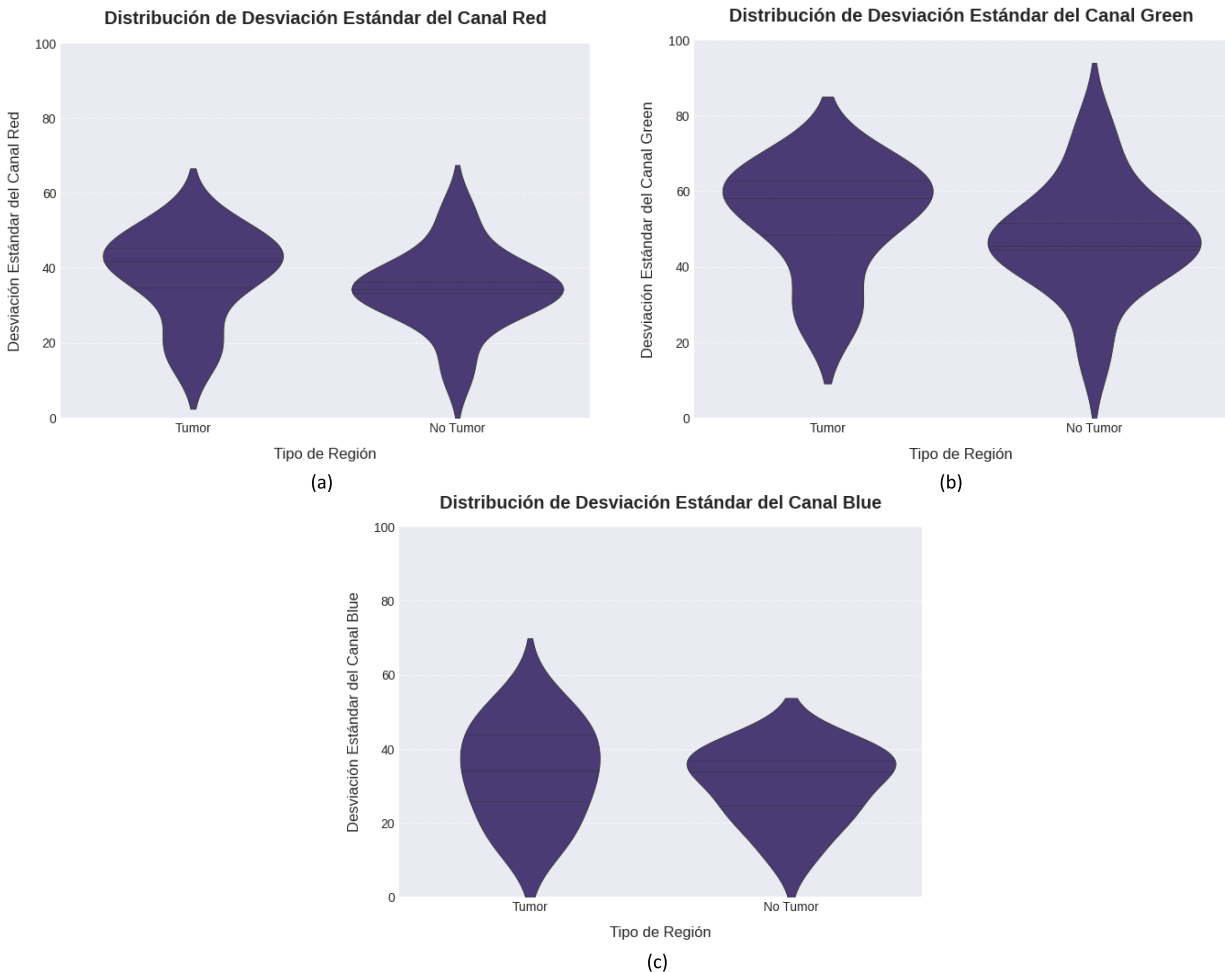


Figura 5.6. Desviación estándar del color en el canal (a) Red, (b) Green y (c) Blue

La identificación y clasificación de parches de tejido inicia con la conversión de la imagen histológica de color (RGB) a escala de grises mediante una transformación lineal ponderada, como $0.299 * R + 0.587 * G + 0.114 * B$. La elección de estos pesos se basa en estándares de procesamiento de imágenes médicas, donde ciertas longitudes de onda asociadas a tinciones de hematoxilina y eosina contribuyen más al contraste entre núcleos y estroma. Para identificar y eliminar regiones no informativas de fondo claro en las imágenes histológicas, se implementa el umbralizado basado en intensidad lumínica, calculando la proporción de píxeles cuya intensidad en escala de grises supera un umbral predefinido $T_1 = 220$, formalizando con $\frac{1}{N} \sum_{i,j} \mathbb{I}(I(i,j) > T_1)$ y la función indicadora \mathbb{I} . Este valor de umbral fue seleccionado experimentalmente para capturar áreas claras que carecen de contenido biológico significativo, asociadas con artefactos de iluminación o zonas de vidrio despejadas.

Para detectar y filtrar imágenes según sus características de fondo se aplican dos criterios principales. El primero consiste en que si la función de detección de píxeles con fondo claro devuelve un valor superior al umbral de 0.98 la imagen se descarta por exceso de claridad de fondo. En segundo lugar, para identificar fondos grises homogéneos se calcula el histograma $H(k)$ de 256 bins y se determina la intensidad pico

k_{peak} definida como $\arg \max_k H(k)$. Si k_{peak} se encuentra entre 120 y 210, y de la misma forma $H(k_{peak})/\sum H(k) > 0.8$ la imagen se clasifica como fondo gris homogéneo y se descarta. Además, a los filtros anteriores, se aplica un criterio adicional para garantizar que la imagen contenga una proporción mínima de tejido informativo. Por esto, se calcula la relación de tejido para determinar el porcentaje de píxeles con intensidad inferior al umbral $T_2 = 120$ considerando que estos valores corresponden a regiones con tejido. La operación se expresa como $\frac{1}{N} \sum_{i,j} \mathbb{I}(I(i,j) < T_2) > 0.01$ y si el resultado de esta operación supera el 1 % la imagen se considera válida; por el contrario, se descarta por insuficiencia de tejido.

Con las imágenes que cumplen con los criterios previos y contienen tejido informativo se procede a segmentar las regiones de interés mediante la umbralización de Otsu para determinar automáticamente el umbral óptimo t^* al maximizar la varianza interclase entre dos grupos. Este enfoque considera ω_1 y ω_2 como las probabilidades de las clases separadas por el umbral t , y μ_1, μ_2 como las medias de intensidad. El umbral óptimo t^* se calcula como $\arg \max_t \left[\omega_1(t)\omega_2(t)(\mu_1(t) - \mu_2(t))^2 \right]$. Sin embargo, en caso de que el histograma de intensidades sea unimodal se requiere al umbral de respaldo basado en el percentil 85 de las intensidades superiores a 30, definido como $P_{85}(I(i,j) | I(i,j) > 30)$.

Finalmente, para el análisis detallado a nivel local, la imagen se divide en una cuadrícula de parches de 256x256 píxeles, donde cada uno es evaluado de forma individual. Cada uno de estos se clasifica como tejido o no tejido según los umbrales previamente presentados, descartando aquellos que no cumplan con los criterios de contenido importante. En el caso de las muestras tumorales, se realiza una correlación con las máscaras de verdad patológica para etiquetar cada parche válido como tumoral o normal.

Posterior a la descripción de la técnica de extracción de parches, la figura 5.7 expone detalladamente el proceso de identificación y selección de tejido en las imágenes WSI del conjunto de datos CAMELYON16. En la sección (a) se presenta un parche representativo de 256x256 píxeles donde se observa claramente las estructuras histológicas incluyendo la disposición celular, morfología de los núcleos y características tisulares clave para el diagnóstico patológico. En forma complementaria, la sección (b) muestra la máscara binaria de segmentación asociada y referenciada como etiqueta de clasificación. En esta, las zonas en blanco representan áreas tumorales, mientras que las zonas en negro corresponden a tejido normal. Por su parte, se realiza una división estratificada del conjunto de datos de 37,800 muestras en tres subconjuntos: entrenamiento con 26,450 (70 %), validación con 5,670 (15 %) y para pruebas 5,670 (15 %). Esta división garantiza consistencia en la distribución de clases y permite reproducibilidad en la división. Inicialmente, se separa el 70 % para entrenamiento y el 30 % se divide en partes iguales para validación y prueba, lo que equivale al 15 % del total en cada uno. A su vez, las proporciones de clases (en este caso 50 % tumorales) se conservan correctamente en todos los conjuntos.

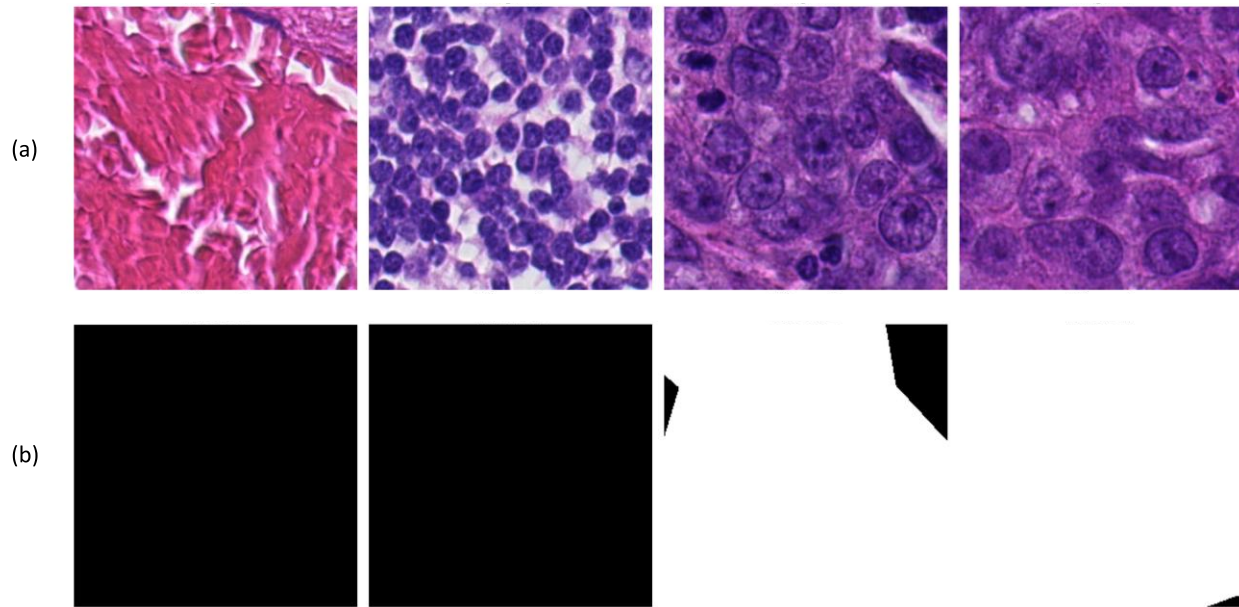


Figura 5.7. Extracción y detección de parches de tejido en las WSI

Para conseguir la consistencia en el análisis de imágenes histológicas se implementa la etapa de normalización de color mediante el método propuesto por Macenko *et al.* [44] que aborda las variaciones cromáticas derivadas de diferencias en la tinción, iluminación o configuración de escáneres. El enfoque se basa en la descomposición espectral de las tinciones de hematoxilina (H) y eosina (E) utilizando un modelo óptico que aplica las leyes de absorción de luz en tejidos teñidos. Inicialmente, las imágenes en espacio RGB se transforman al espacio de densidad óptica (OD) mediante la ley de Beer-Lambert, expresada como $-\log I$, donde I representa el vector de intensidades RGB normalizados. Para completar el proceso de normalización de color, se aplica la descomposición en valores singulares (SVD) a la matriz de densidad óptica con el objetivo de estimar los vectores de tinción que definen las direcciones cromáticas dominantes en la imagen histológica.

Este procedimiento, representado como $S = V^{-1} * OD$, modela la interacción entre la absorción de luz y los colorantes utilizados en la tinción H&E. En la ecuación, V contiene los vectores de tinción que caracterizan las componentes principales de color. Esta descomposición permite separar y cuantificar la contribución individual de cada tinción para facilitar la estandarización de intensidades. Para finalizar el proceso de normalización cromática, se realiza una proyección y ajuste lineal de las proyecciones de las intensidades de color. En esta etapa, las intensidades originales de la imagen se proyectan sobre los vectores de tinción estimados previamente, lo que permite separar y cuantificar la contribución individual de cada colorante en la muestra histológica. Por otra parte, estas intensidades se normalizan mediante una transformación lineal utilizando la media y la desviación estándar de las distribuciones de intensidad de las tinciones, preservando las características morfológicas del tejido mientras se eliminan artefactos causados por diferencias en la adquisición de las muestras. Por último, la imagen normalizada se reconstruye en el espacio RGB mediante la combinación de las intensidades ajustadas con los vectores de tinción, lo que da como resultado una representación visualmente consistente, asegurando que las características extraídas sean robustas y representativas de las propiedades biológicas de la imagen.

En la figura 5.8 se ilustran los resultados obtenidos mediante la aplicación del método de normalización de color propuesto por Macenko *et al.* a los parches de imágenes histológicas. Para este proceso, se seleccionó una imagen de referencia (a) que exhibe una tinción H&E óptima que se caracteriza por un equilibrio entre la hematoxilina y la eosina. Esta imagen de referencia sirvió como patrón para ajustar las características cromáticas de todo el conjunto de datos y asegurar la estandarización de las intensidades de color en las muestras procesadas. En la sección (b) se presentan cinco parches de prueba, seleccionados estratégicamente para representar a variabilidad común encontrada en las preparaciones histológicas, como diferencias en la intensidad de tinción, niveles de contraste y tonos de color. Esta cantidad de parches fue elegida para proporcionar una evaluación robusta del método de normalización, permitiendo capturar una visión amplia de variaciones técnicas sin llegar a ser redundante. Como resultado, se observa en (c) una distribución equilibrada y uniforme de los colorantes en las imágenes normalizadas, donde los núcleos celulares y las estructuras de tejido se distinguen con mayor claridad y consistencia. Por lo consiguiente, el método garantiza que las diferencias observadas en las imágenes representen verdaderas variaciones biológicas y no artefactos técnicos, optimizando la validez de los hallazgos en el diagnóstico histológico.

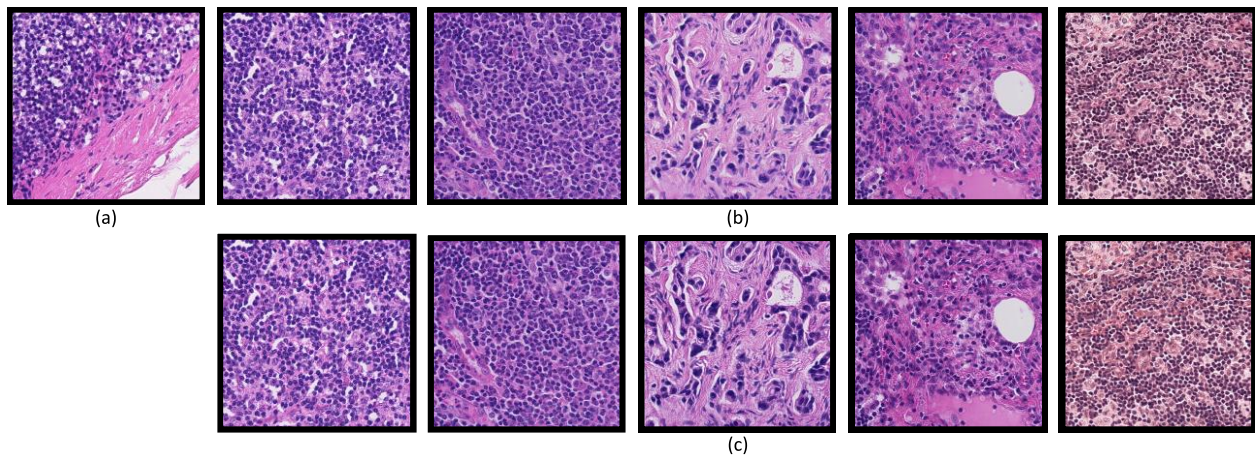


Figura 5.8. Normalización de color del conjunto de datos. (a) Imagen de referencia, (b) parches de prueba y (c) Distribución equilibrada de colorantes

Para cuantificar la eficacia del método de normalización de Macenko sobre los cinco parches de prueba de imágenes H&E, se empleó un enfoque de evaluación dual mediante los descriptores ORB (Oriented FAST and Rotate BRIEF) y el índice SSIM (Structural Similarity Index Measure). La métrica ORB evalúa la preservación de la información estructural clave, como núcleos celulares y bordes tisulares antes y después de la normalización, midiendo concretamente la cantidad de puntos de interés coincidentes entre las imágenes para verificar que la transformación cromática no alterara la integridad morfológica [45]. En complemento, el índice SSIM proporcionó una evaluación perceptual de la similitud estructural global, considerando factores como luminancia, contraste y textura, para determinar si la información anatómica se mantuvo intacta y sin distorsiones significativas [46]. Los gráficos de violín mostrados en la figura 5.9 que presentaron morfología de embudo evidencian una distribución de similitud donde la mayoría de los parches presentan valores altos y concentrados (cerca de 1.0), evidenciando por la parte ancha del embudo en la zona superior, lo que indica una correspondencia casi perfecta en los métodos ORB y SSIM para los parches 1, 2, 4 y 5. Sin embargo, la base estrecha y alargada del gráfico hacia valores más bajos

está impulsada únicamente por el parche 3, cuyas similitudes (0.889 en ORB y 0.862 en SSIM) son notablemente inferiores, actuando como un *outlier*. Esta forma característica revela que, si bien el proceso de comparación es generalmente efectivo y consistente, el parche 3 representa una anomalía específica donde la similitud se reduce significativamente, posiblemente debido a una variación local en la textura, iluminación o contenido que ambos métodos logran detectar de forma consistente.

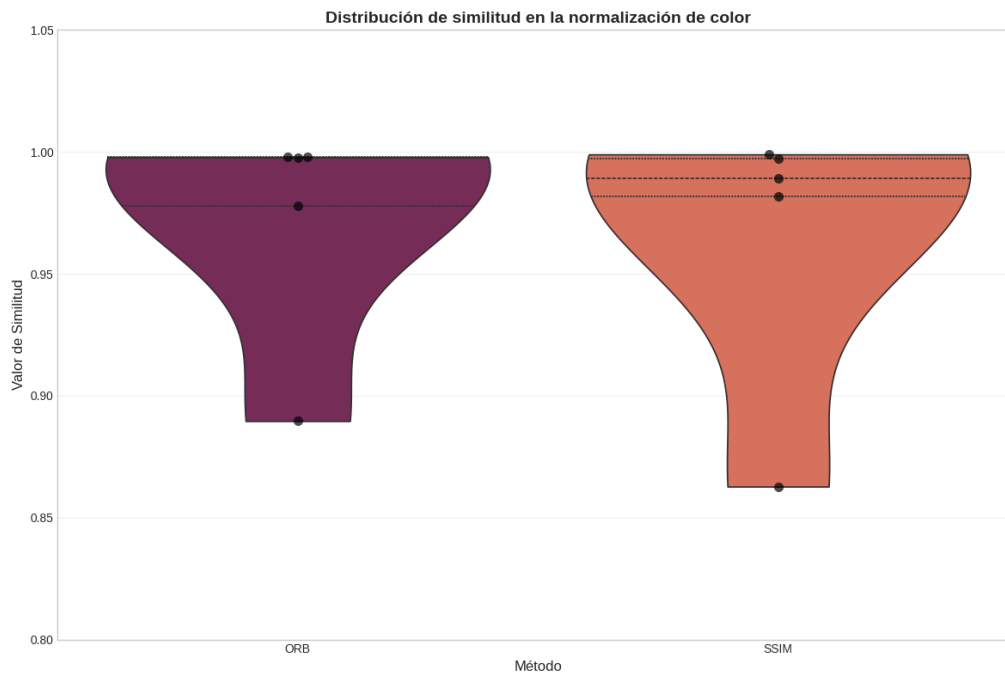


Figura 5.9. Distribución de similitud en la normalización de color

Para enriquecer el conjunto de datos de entrenamiento y mejorar la generalización de los modelos se implementa un proceso de aumento de datos. Este enfoque aplica un conjunto de transformaciones que amplía la diversidad de los datos mientras se preserva la validez anatómica y las características clínicamente relevantes de las imágenes histológicas. Entre las transformaciones aplicadas se incluyen rotaciones, volteos, ajustes de brillo y contraste incluyendo variaciones de color, todas estas calibradas para mantener la integridad diagnóstica de las estructuras tisulares. De la misma forma, el aumento de datos incorpora variaciones simuladas en condiciones de adquisición, relacionadas con diferencias de iluminación, parámetros de tinción y configuración de equipos, lo que genera robustez frente a la heterogeneidad encontrada en entornos reales. Estas modificaciones permiten que los modelos aprendan a reconocer patrones histológicos independientes de las variaciones técnicas, mejorando la capacidad de generalización. La tabla 2 resume detalladamente los parámetros aplicados, organizados por categorías, junto con sus rangos de variación.

Tabla 2. Transformaciones para el aumento de datos

Categoría	Transformación	Parámetros	Probabilidad
Transformaciones Geométricas (probabilidad)	Rotación	Máximo 15° de inclinación	50 % (dentro de la categoría)

global: 70 %	Escalado afín	Ajuste de tamaño entre 90 % y 110 %	-
	Deformación elástica	Intensidad baja ($\alpha=0.5$), Suavizado moderado ($\sigma=15$), Deformación elástica suave ($\alpha_{\text{affine}}=5$)	20 % (dentro de la categoría)
Volteos	Volteo horizontal	-	50 %
	Volteo vertical	-	50 %
Ajuste de color (Probabilidad global: 50%)	Variación de color	Brillo (10%), contraste (10%), saturación (10%), tono (5%)	30 % (dentro de la categoría)
	Modificación de tono y saturación	Rango de tono (± 10), saturación (± 15), valor (± 10)	30 % (dentro de la categoría)
Ruido y Desenfoque (Probabilidad global: 40 %)	Desenfoque Gaussiano	Tamaño de kernel entre 1x1 y 3x3 píxeles	30 % (dentro de la categoría)
	Ruido Gaussiano	Intensidad de ruido entre 5.0 y 20.0 de varianza	20 % (dentro de la categoría)

La figura 5.10 presenta detalladamente la etapa de aumentado de datos implementado para enriquecer el conjunto de entrenamiento donde se aplican las transformaciones seleccionadas tanto a los parches de tejido (a) como a sus máscaras de segmentación (b). Cada transformación previamente mencionada se aplica de manera sincronizada a las imágenes y sus máscaras para garantizar la correspondencia espacial entre los parches aumentados y sus transformaciones. Este enfoque mejora la robustez del modelo con respecto a variaciones técnicas y prepara al sistema para generalizar apropiadamente datos reales con condiciones heterogéneas.

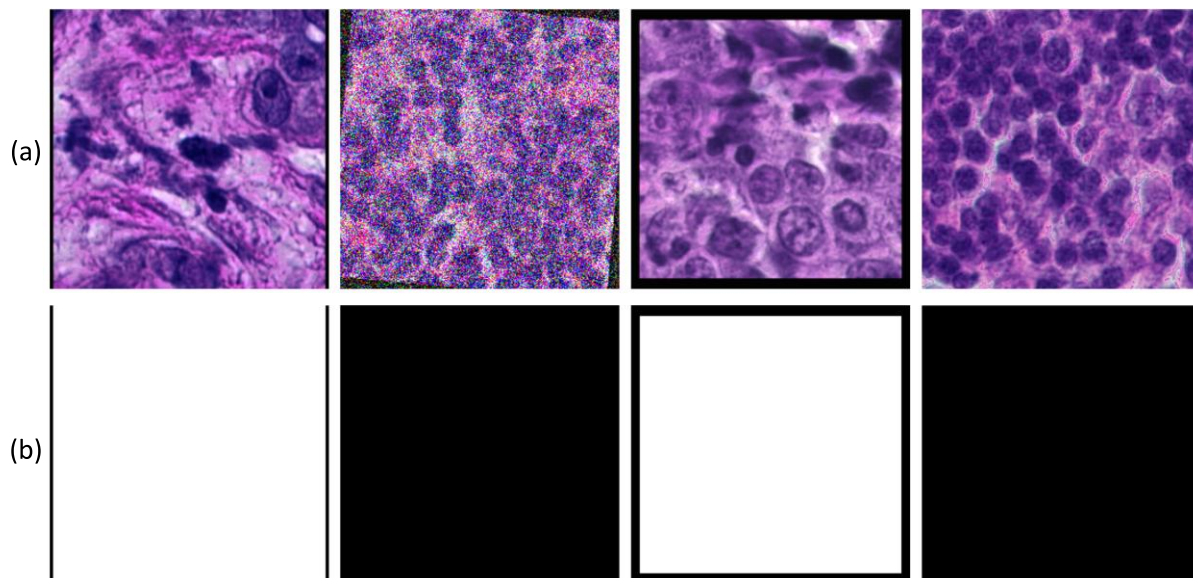


Figura 5.10. Aumentado de datos para muestras de entrenamiento. (a) Parches de tejido y (b) Máscaras de segmentación

6. ESTIMACIÓN DE CARGA TUMORAL MEDIANTE MODELOS DE SEGMENTACIÓN DE APRENDIZAJE MÁQUINA

Para el entrenamiento y desarrollo de las arquitecturas de segmentación de imágenes histológicas H&E de cáncer de mama, se utilizó el entorno de computación acelerada proporcionada por Google Colab. Esta plataforma ofreció acceso a una GPU NVIDIA T4, cuyo hardware especializado fue fundamental para manejar la alta demanda computacional de los modelos de red neuronal profunda. La arquitectura NVIDIA Turing de esta GPU, junto con sus 16 GB de memoria GDDR6 y un amplio ancho de banda de hasta 320 GB/s, permitió procesar eficientemente los grandes datos de volúmenes de imágenes de alta resolución, facilitando un flujo de trabajo ágil durante el entrenamiento. El rendimiento computacional estuvo particularmente potenciado por los Tensor Cores de segunda generación, núcleos diseñados específicamente para operaciones de precisión mixta (FP16/FP32). Esta optimización aceleró de forma significativa tanto las fases de entrenamiento como de inferencia de los modelos de segmentación. Este recurso se complementó con la configuración estándar del entorno de Google Colab, que asignó aproximadamente 12.7 GB de RAM de sistema junto con la GPU. Esta combinación fue suficiente para gestionar en memoria las operaciones de preprocesamiento, el manejo de los conjuntos de datos de imágenes histológicas y las iteraciones del entrenamiento. Las versiones específicas de las bibliotecas principales empleadas en este entorno se detallan en la tabla 3. Para garantizar la reproducibilidad de todos los resultados, se fijó una semilla aleatoria (random seed) en 42, utilizada en forma sistemática tanto en la división de los conjuntos de datos como en la iniciación de los pesos de los modelos.

Tabla 3. Versiones específicas de las bibliotecas principales

Biblioteca	Versión	Descripción
Python	3.12.12	Lenguaje de programación de alto nivel
NumPy	2.0.2	Biblioteca de Python para el cálculo matricial
Pandas	2.2.2	Biblioteca utilizada para la manipulación y análisis de datos
Matplotlib	3.10.0	Biblioteca para crear visualizaciones de datos de alta calidad
OpenSlide	1.4.3	Biblioteca que proporciona una interfaz para leer imágenes de portaobjetos completos de patología digital
TensorFlow	2.19.0	Librería para el desarrollo de modelos profundos
OpenCV	4.12.0	Librería especializada de visión por computador
Scikit-image	0.25.2	Paquete de aprendizaje automático para el procesamiento de imágenes

Una vez definida la proporción de datos para entrenamiento y prueba, se propone una arquitectura híbrida UNet optimizada con ResNet50 como codificador que se muestra en la figura 6.1. En esta arquitectura, se congelan las primeras 100 capas y se aplica regularización L2 con un valor de 0.00001 en las capas no congeladas. El codificador ajustado incluye bloques de muestreo hacia arriba que emplean convoluciones transpuestas inicializadas con el método He normal, seguido de normalización por lotes, activación LeakyReLU con parámetro alpha de 0.1 y Dropout adaptativo que se aplica con una tasa de 0.4 cuando las características son mayores o iguales a 128, y de 0.2 en los demás casos. Las conexiones directas (skip connections) unen estratégicamente los mapas de características del codificador, extraídos de las capas conv1_relu, conv2_block3_out y conv3_block4_out de ResNet50 siguiendo los niveles del decodificador. Para regularización de estas conexiones se emplea *dropout* espacial con tasas decrecientes de 0.3, 0.2 y 0.1 según la profundidad de la capa. Por otro lado, el puente incorpora convoluciones dilatadas con una

tasa de 2, regularización L2 de 0.0001 y *dropout* espacial de 0.4 para ampliar el campo receptivo. Finalmente, en la última capa se utiliza la función de activación softmax inicializada con el método Glorot Uniform junto con regularización L2 de 0.00001.

El modelo emplea una jerarquía de canales que comienza con 64 canales en mapas de características de 128x128 que se incrementa progresivamente a 256 canales (64x64), 512 canales (32x32) y alcanza su máxima profundidad en el codificador con 1024 canales (16x16). A su vez, el módulo central procesa estos datos con la reducción a 512 canales, mientras que el decodificador disminuye gradualmente este número mediante el proceso de aumento de resolución. En términos de capacidad, el codificador concentra la mayor parte de los parámetros con 8,589,184 que equivale al 55 % del total, seguido por el módulo central con 4,719,104 parámetros y el decodificador con 2,324,832 parámetros, destacando la capa final con su eficiencia con solo 66 parámetros. Con un total de 15,635,234 parámetros, 11,520,994 ajustables durante el entrenamiento, la arquitectura logra un equilibrio entre potencia de representación y eficiencia computacional debido a la distribución que prioriza el codificador como núcleo que procesa características.

El modelo se configura utilizando el optimizador AdamW con una tasa de aprendizaje inicial de 0.0001 y un decaimiento de pesos de 0.0001, completando con función de pérdida focal binaria ajustada con un gamma de 2.0 y suavizado de etiquetas del 5 %. Durante el entrenamiento, se implementan mecanismos de control que incluyen parada temprana con paciencia de 10 épocas, reducción automática de la tasa de aprendizaje con un factor de 0.1 después de 10 épocas sin mejora y almacenamiento del mejor modelo según el menor valor de pérdida en validación. El proceso se desarrolla con lotes de 4 muestras por actualización de pesos, durante 50 iteraciones completas sobre los datos, procesando 200 lotes de entrenamiento y evaluando 50 lotes de validación en cada ciclo, optimizando de esta forma el equilibrio entre eficiencia computacional y capacidad de aprendizaje.

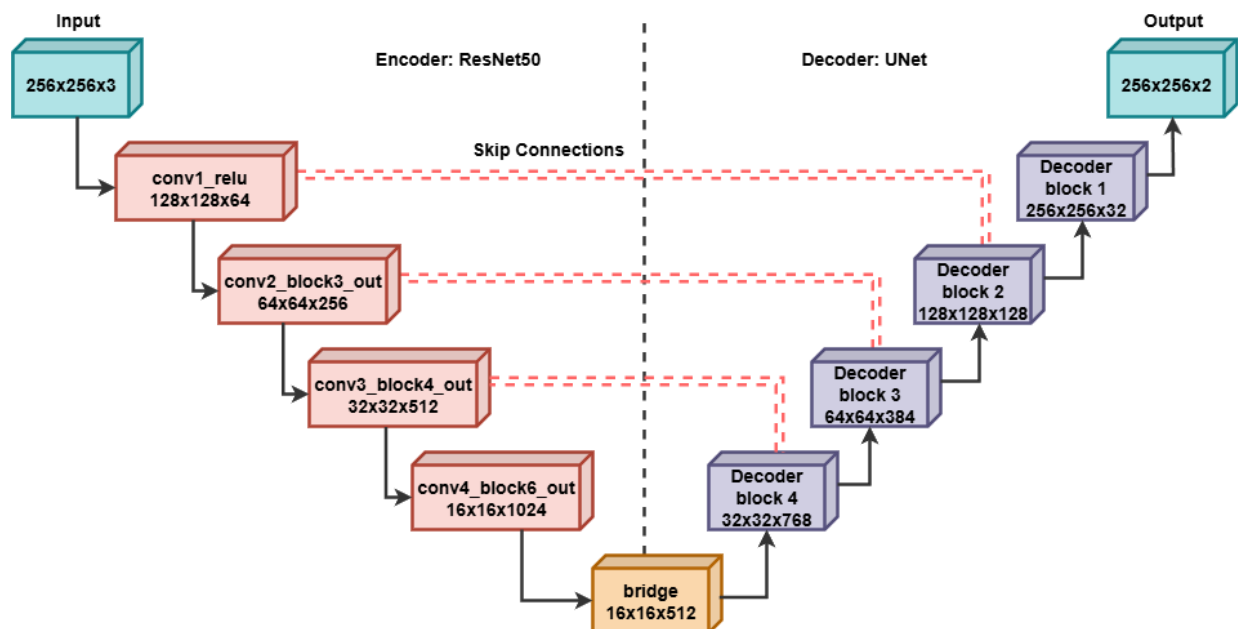


Figura 6.1. Arquitectura UNet + ResNet50 desarrollada

Conociendo la estructura híbrida UNet optimizada con ResNet50 posteriormente se entrenó el modelo. Se

mostró una evolución significativa durante las 50 épocas, demostrando un aprendizaje efectivo con una mejora consistente tanto en la pérdida (loss) como en la exactitud (accuracy) en los conjuntos de entrenamiento y validación. En primer lugar, como se presenta en la figura 6.2, la pérdida de entrenamiento disminuyó de 0.6841 a 0.1050, mientras que la de validación cae de 0.4084 a 0.0647, lo que señala una reducción significativa del error y capacidad de generalización.

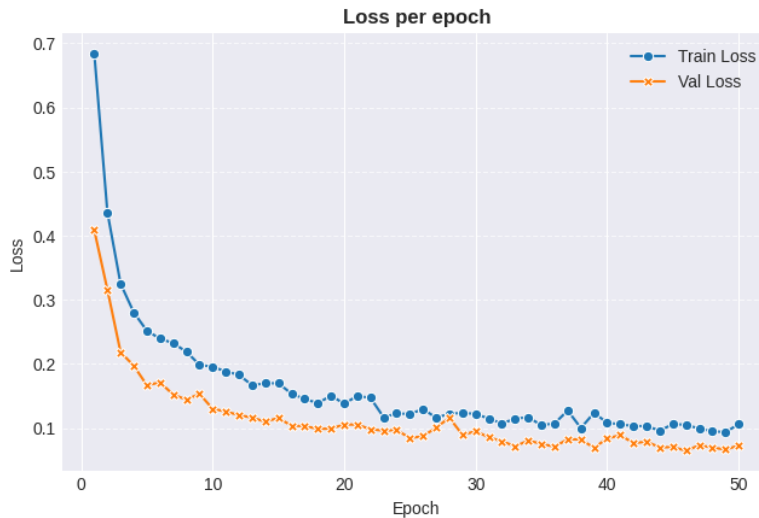


Figura 6.2. Evolución de la pérdida del modelo UNet + ResNet50

La exactitud de entrenamiento incrementó de 0.5427 a 0.9111, y en validación cambia de 0.6484 a 0.9628, registrando una capacidad de diferenciación correcta. El rendimiento general es consistente, con alta exactitud y error bajo en validación, demostrando un buen equilibrio entre sesgo y varianza. En la figura 6.3 se presenta la evolución en la métrica de exactitud. El modelo híbrido demostró una notable eficiencia computacional, completando su fase de entrenamiento en 31.98 minutos, un periodo que subraya la efectividad del diseño propuesto y su potencial de ser implementado en escenarios que requieren rápidas iteraciones.

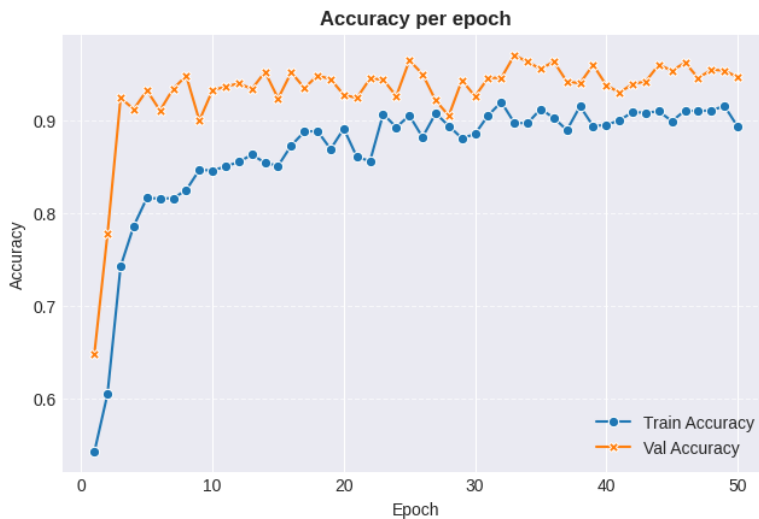


Figura 6.3. Evolución de la exactitud del modelo UNet + ResNet50

Similarmente, el modelo presentó un rendimiento notable en las métricas de AUC (Área bajo la curva) y precisión, evidenciando una fuerte capacidad de discriminar entre las clases y clasificar efectivamente las instancias verdaderas. El AUC para validación experimenta una mejora, pasando de 0.7357 en la primera época a un máximo de 0.9961, denotando una excelente separación entre las clases y bajo solapamiento en las distribuciones de probabilidades predichas.

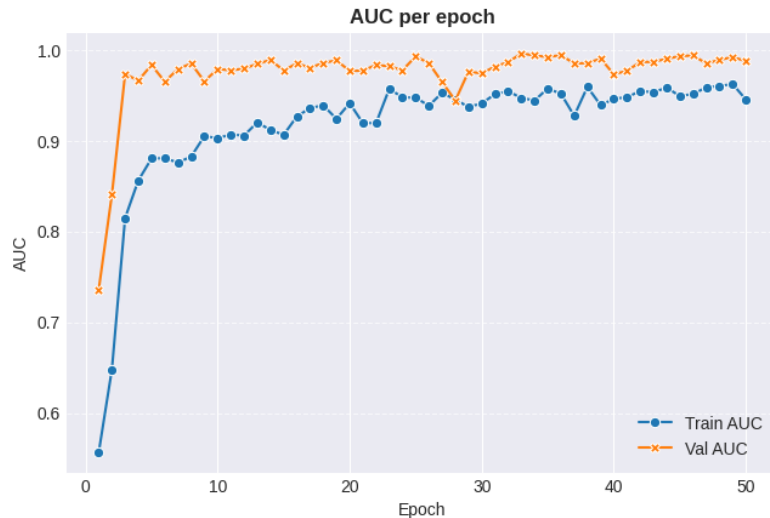


Figura 6.4. Evolución del AUC del modelo UNet + ResNet50

Por otro lado, la precisión en validación también muestra un notable progreso, aumentando de 0.6484 a 0.9708, lo que refleja una alta fiabilidad en las predicciones positivas del modelo. No obstante, se observan fluctuaciones ligeras en las últimas épocas, lo que sugiere un pequeño sobreajuste. Aún así, los valores máximos confirman que el modelo logra un equilibrio robusto entre sensibilidad y especificidad con una capacidad predictiva alta.

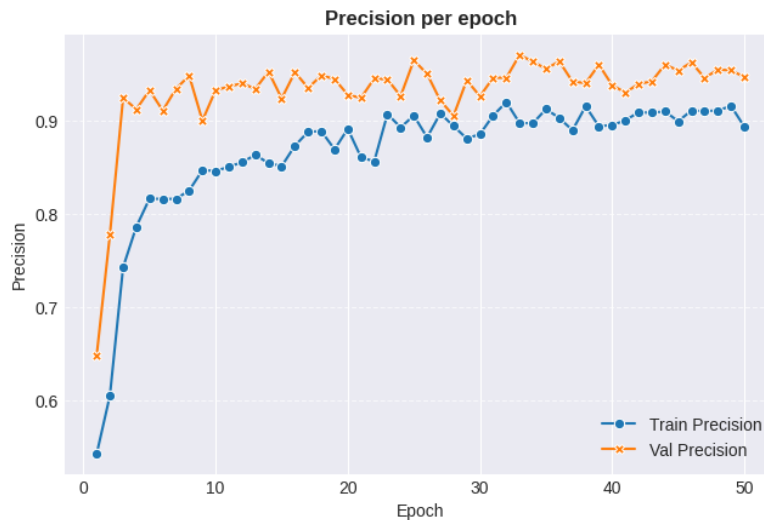


Figura 6.5. Evolución de la precisión del modelo UNet + ResNet50

El modelo presentó un *recall* (sensibilidad) en validación que evoluciona de forma consistente (figura 6.6),

partiendo de 0.6484 en la primera época hasta lograr un valor máximo de 0.9708 en la época 33, demostrando una capacidad de detectar eficientemente las instancias positivas. Esta mejoría, junto con la precisión alta y AUC, demuestran que el modelo no solo presenta robustez, sino también sensible a las clases relevantes, minimizando la influencia de falsos negativos.

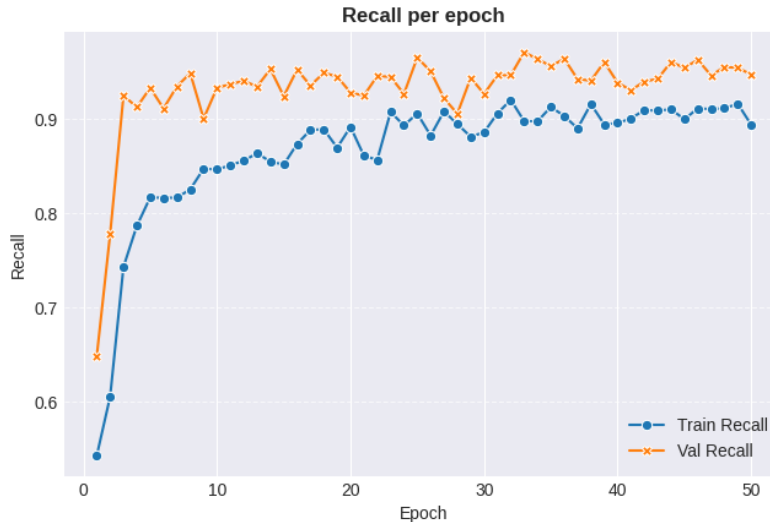


Figura 6.6. Evolución del *recall* del modelo UNet + ResNet50

Respecto al tiempo de ejecución por paso (ms/step) presentado en la figura 6.7 se aprecia una variabilidad representativa, desde 157 ms/step en el primer paso hasta 228 ms/step en épocas posteriores, con un promedio de 180 ms/step. Estas fluctuaciones pueden ser causa a procesos en segundo plano, gestión de memoria y variaciones del sistema, pero no representan un problema en el modelo. De todos modos, los tiempos se mantienen en un rango moderado para un entrenamiento de aprendizaje profundo sin evidencia embotellamientos.

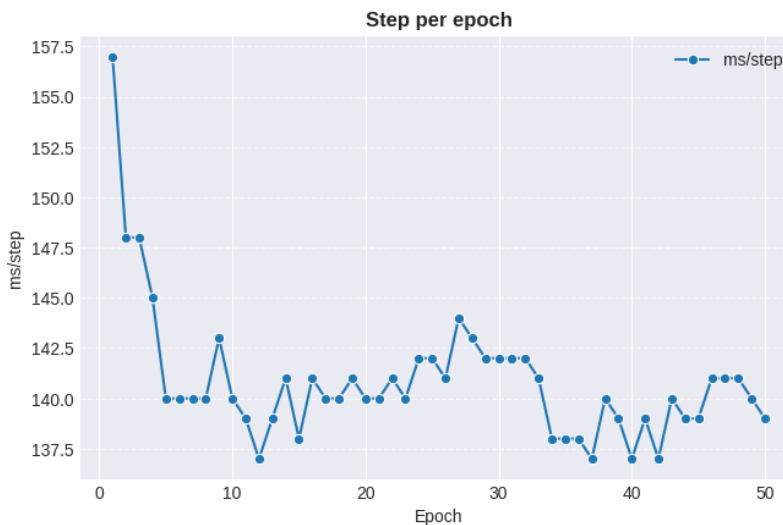


Figura 6.7. Evolución del *step* por época del modelo UNet + ResNet50

Para validar y consolidar el desempeño del modelo híbrido UNet + ResNet50 propuesto para la

segmentación de ganglios linfáticos y estimación de carga tumoral en cáncer de mama, se desarrolló un análisis comparativo con las arquitecturas ampliamente utilizadas de aprendizaje profundo ResNet50, UNet y DenseNet121. Para garantizar una comparación imparcial y reproducible entre los diferentes modelos, todos fueron evaluados bajo las mismas condiciones experimentales utilizando los mismos hiperparámetros y configuraciones dentro del algoritmo híbrido implementado, los cuales se detallan en la tabla 4. Con este enfoque, se eliminan posibles sesgos y se asegura que las diferencias observadas en el desempeño de los modelos se atribuyan a sus características intrínsecas y no a variaciones en los parámetros de evaluación.

Tabla 4. Hiperparámetros aplicados a cada modelo

Categoría	Hiperparámetro	Valor
Optimización	Optimizador	AdamW
	Tasa de aprendizaje inicial (Learning Rate)	0.0001
	Decaimiento de pesos (Weight Decay)	0.0001
Pérdida	Función de pérdida	Focal Loss Binaria
	Gamma	2.0
	Suavizado de etiquetas (Label Smoothing)	5 %
Entrenamiento	Épocas (Epochs)	50
	Tamaño de lote (Batch Size)	4
	Lotes de entrenamiento por época	200
	Lotes de validación por época	50
Control de entrenamiento	Parada temprana (Early Stopping)	Paciencia: 10 épocas
	Reducción de LR (Learning Rate)	Factor: 0.1
	Monitorización	Pérdida de validación
	Paciencia para reducción de LR	10

Con respecto al modelo de segmentación presentado en la figura 6.8 se utiliza una arquitectura basada en codificador-decodificador con conexiones laterales. La parte codificadora se basa en la red ResNet50 preentrenada con las primeras 100 capas bloqueadas para extraer características en cuatro escalas. La primera es una capa convolucional de 128x128 con 64 filtros, la siguiente es la salida del segundo bloque de 64x64 con 256 filtros, la tercera consiste en la salida del bloque 3 a 32x32 con 512 filtros y la última es la salida del cuarto bloque de 16x16 con 1024 filtros, usando la salida del quinto bloque de 8x8 con 2048 filtros como capa central. Por otro lado, la parte decodificadora amplía progresivamente la resolución mediante capas transpuestas de convolución con 512, 256, 128 y 64 filtros, combinadas con las características de la codificadora, cada una seguida por operaciones de convolución 3x3, normalización por lotes y anulación aleatoria de neuronas con 30 % en los bloques iniciales y 20 % en los últimos. La capa final utiliza una convolución transpuesta de 32 filtros para asegurar el tamaño original de la imagen, generando una salida binaria a partir de una convolución 1x1. El modelo presenta una arquitectura con un total de 43,970,306 parámetros, distribuidos entre el codificador y decodificador. En el codificador, se emplean 23,587,712 parámetros fijos, los cuales corresponden a capas preentrenadas y no ajustables durante el entrenamiento, lo que permite aprovechar las representaciones ya aprendidas. Por otro lado, el decodificador cuenta con 20,382,594 parámetros ajustables, los cuales son optimizados durante el entrenamiento para adaptación en la segmentación, permitiendo mayor flexibilidad y capacidad de aprendizaje.

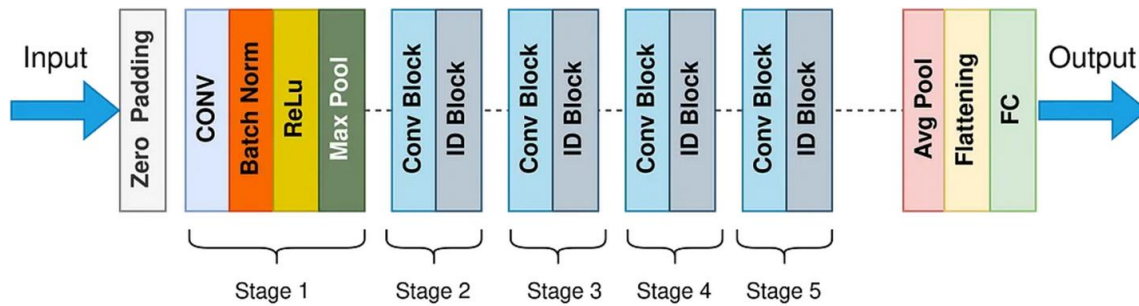


Figura 6.8. Arquitectura ResNet50. Tomada de [47]

La evolución en el entrenamiento del modelo inicialmente mostró un buen rendimiento con una mejora en las métricas en las primeras épocas. Como se muestra en la figura 6.9, la pérdida de entrenamiento disminuyó desde 0.2273 en la primera época hasta 0.0979 en la época 19, mientras que el *accuracy* de entrenamiento aumenta del 62.88 % al 85.89 % indicando aprendizaje efectivo tal como se presenta en la figura 6.10.



Figura 6.9. Evolución de la pérdida del modelo ResNet50

No obstante, la pérdida de validación presentó mejoría en las primeras épocas de 0.08411 a 0.04808 en la época 9, pero luego fluctúa, indicando la presencia de sobreajuste, debido a que el *accuracy* de entrenamiento sigue mejorando mientras que en validación se estanca entre 95 %. La parada temprana en la época 19 activado después de reducir la tasa de aprendizaje de 0.0001 a 0.00001 evita un sobreentrenamiento.

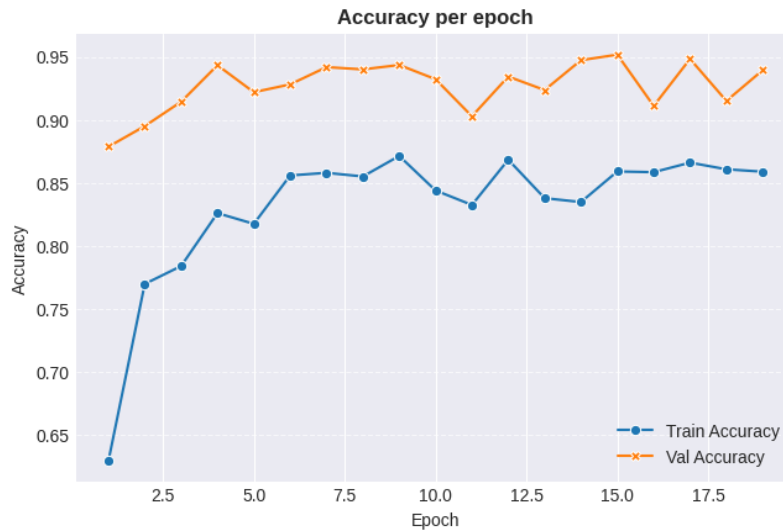


Figura 6.10. Evolución de la exactitud del modelo ResNet50

En general, el modelo logró una exactitud en validación de 94.36 % en la época 9, pero presentó alta divergencia entre las curvas de entrenamiento y validación, afectando la generalización. El modelo completó su entrenamiento en un tiempo de 26.82 minutos. Sin embargo, a pesar de esta eficiencia computacional, se evidenciaron signos de sobreajuste, indicando que el modelo aprendió detalles específicos del conjunto de entrenamiento en lugar de generalizar patrones de los datos no vistos.

Como se aprecia en la figura 6.11, el modelo demuestra un desempeño discriminativo alcanzando un AUC de validación máximo de 0.9884 en la época 7 y una precisión de validación de 94.58 % en la época 9, indicando capacidad de minimizar falsos positivos y distinguir entre clases. Sin embargo, se observó una divergencia significativa entre las métricas de entrenamiento y validación. El AUC de entrenamiento fluctúa entre 0.8370 en la época 3 y 0.9374 en la época 17, mientras que el AUC en validación se mantiene en el rango de 0.9408 a 0.9884, indicando que el modelo aprende patrones ruidosos en etapas intermedias.

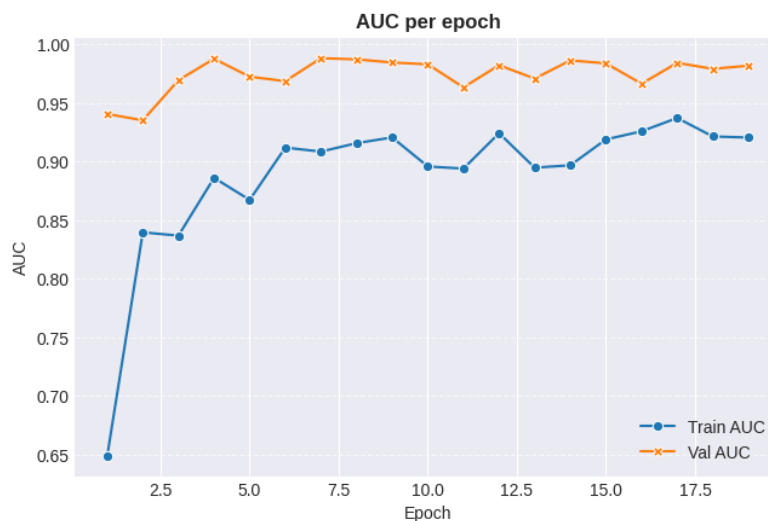


Figura 6.11. Evolución del AUC del modelo ResNet50

La precisión en entrenamiento, que osciló entre 85.59 % y 87.14 % presenta un margen de 7 a 9 puntos porcentuales inferior a la precisión en validación que varía entre 91.11 % y 95.18 %, evidenciando la presencia de sobreajuste. Esta diferencia indica que el modelo se ajusta a los datos de entrenamiento, capturando no solo patrones generales, sino también ruido y peculiaridades específicas de los datos. A su vez, la parada temprana en la época 19, activada después de reducir la tasa de aprendizaje a 0.00001 evitó un sobreentrenamiento.

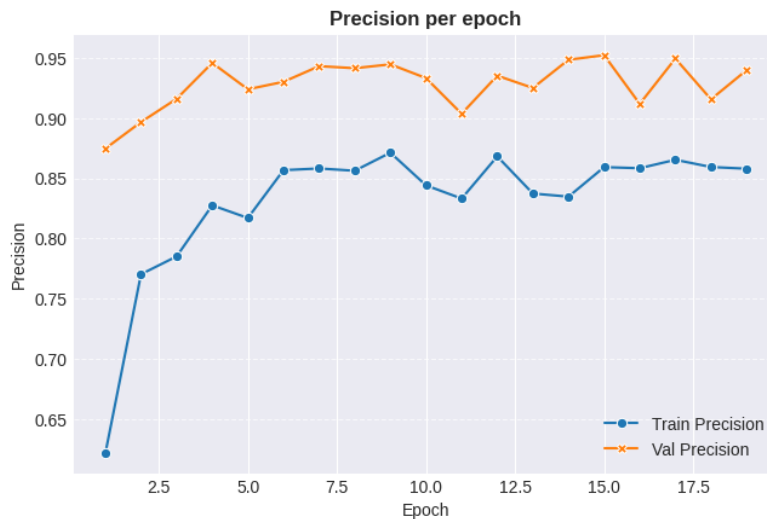


Figura 6.12. Evolución de la precisión del modelo ResNet50

Por otra parte, el modelo muestra un *recall* de validación que inicia en 87.74 % en la época 1, alcanzando 95.08 % en la época 15 que supera de forma constante el *recall* de entrenamiento, evolucionando desde 60.20 % hasta 86.60 %. Esta diferencia de 8.14 a 9.48 puntos porcentuales entre los conjuntos de entrenamiento y validación sugiere que, aunque el modelo identifica los casos positivos no vistos podría presentar insensibilidad a patrones específicos presentes en los datos de entrenamiento. La evolución en la métrica *recall* se presenta en la figura 6.13.

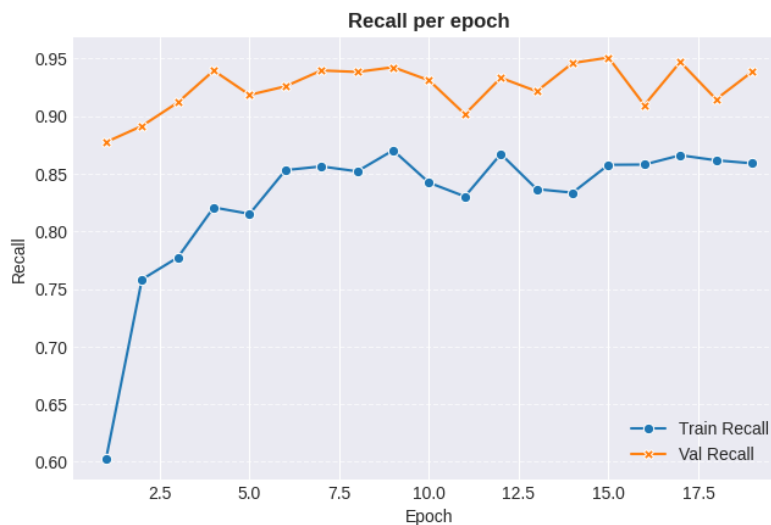


Figura 6.13. Evolución del *recall* del modelo ResNet50

En cuanto al rendimiento computacional, el tiempo por paso muestra fluctuaciones, iniciando en 229 ms/step y alcanzando un máximo de 673 ms/step en la época 4, luego se estabilizó entre 260 y 400 ms/step en épocas posteriores, con un valor final de 393 ms/step en la época 19, con variaciones ocasionadas por operaciones de guardado de modelos y carga computacional variable.

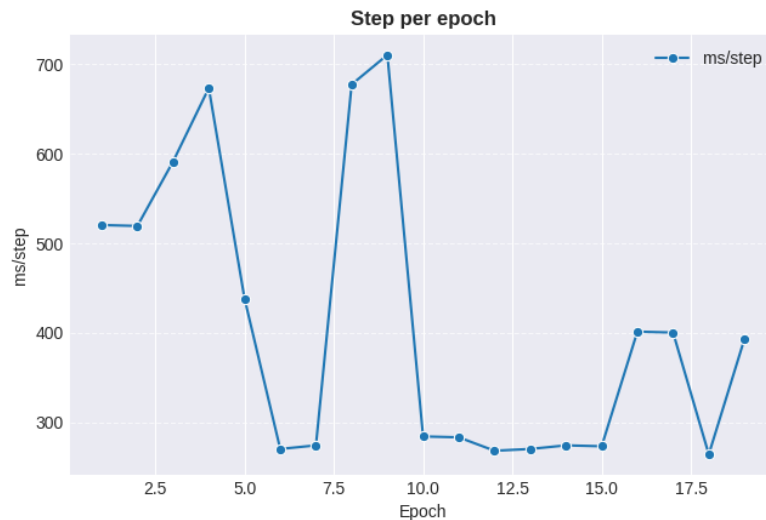


Figura 6.14. Evolución del *step* por época del modelo ResNet50

El modelo UNet implementado y mostrado en la figura 6.15 sigue una estructura simétrica con una red codificadora que reduce progresivamente el tamaño de los datos y una red decodificadora que los reconstruye, conectadas mediante enlaces directos entre bloques correspondientes. La parte codificadora consta de 4 etapas. La primera etapa aplica dos convoluciones de 64 filtros de tamaño 3x3 con función de activación ReLU y normalización por lotes, seguido de normalización de tamaño mediante MaxPooling de tamaño 2x2. Este proceso se repite para la segunda etapa con 128 filtros, tercera etapa con 256 filtros y la cuarta etapa con 512 filtros.

Por otra parte, la etapa central del modelo utiliza 1024 filtros y la parte codificadora realiza el proceso inverso mediante convoluciones transpuestas para aumentar el tamaño con 512, 256, 128 y 64 filtros, combinando estos resultados con la información almacenada en la codificadora por medio de concatenación y aplicación de convoluciones. La salida genera la máscara de segmentación con dos canales usando la función sigmoide, alcanzando un total de 31,055,362 parámetros ajustables, distribuidos en 18,876,160 parámetros en la parte que reduce la imagen (codificador) y 12,179,202 parámetros en la parte que aumenta la imagen (decodificador). La capa final tiene 130 parámetros y se presentan 11,776 parámetros no entrenables.

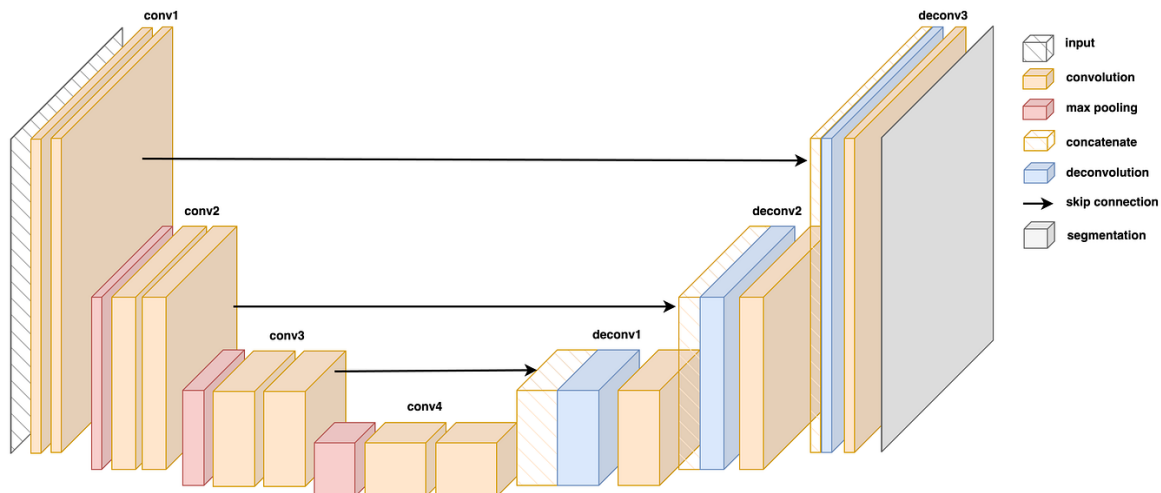


Figura 6.15. Arquitectura U-NET. Tomada de [48]

El análisis de las métricas de pérdida (figura 6.16) y exactitud (figura 6.17) revela un proceso de entrenamiento con una tendencia positiva. Sin embargo, se marcan notables fluctuaciones y signos de sobreajuste, donde si bien la pérdida de validación disminuye desde 0.1679 hasta un mínimo de 0.04475 en la época 22, esta métrica muestra una inestabilidad pronunciada con repentinos incrementos, como el pico de 0.1419 en la época 4 y 0.1088 en la época 26 que delatan una capacidad inconsistente de generalización que se valida con las oscilaciones en la exactitud de validación que llega a caer desde 93.96 % en la época 13 hasta 87.55 % en la época 14, y hasta 82.30 % en la 26, evidenciando una brecha variable y amplia con la exactitud de entrenamiento que refleja momentos donde el modelo se ajusta a los datos de entrenamiento.

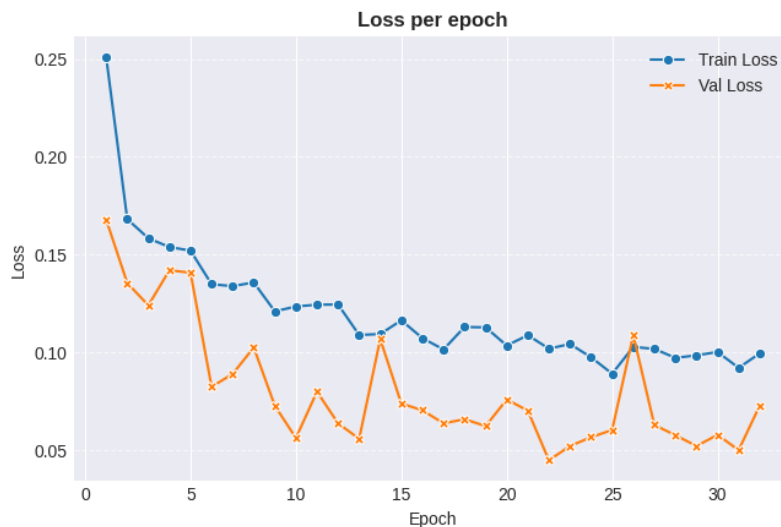


Figura 6.16. Evolución de la pérdida del modelo UNet

Finalmente, con la estrategia de reducción de la tasa de aprendizaje y el *early stopping* se alcanza un rendimiento de 94.25 % de exactitud en validación en la época 22. A pesar de que el modelo demostró una eficiencia computacional al entrenarse en 46.08 minutos, su utilidad se vio comprometida por presentar

sobreajuste. Este fenómeno en un modelo que, si bien aprendió a la perfección los ejemplos de entrenamiento, mostró una deficiente capacidad para generalizar y realizar predicciones sobre los datos nuevos no vistos.

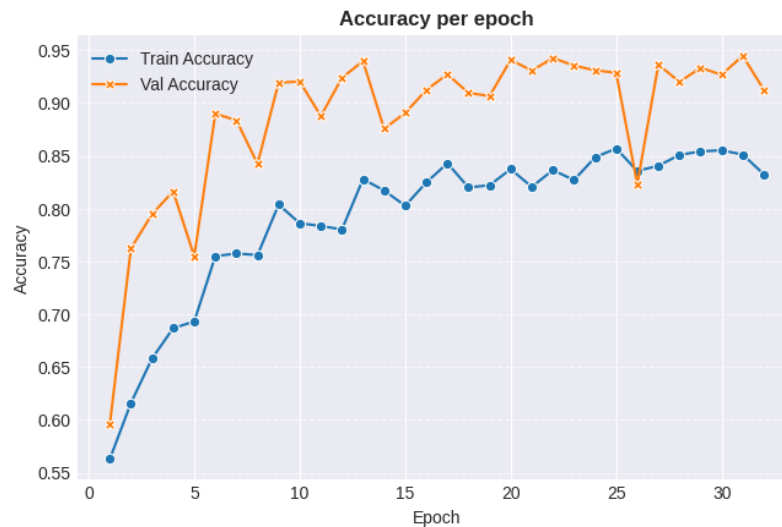


Figura 6.17. Evolución de la exactitud del modelo UNet

El análisis de las métricas de AUC y precisión revelan un desempeño con tendencia positiva. No obstante, como se aprecia en la figura 6.18, se encuentra marcado por significativas fluctuaciones y evidencias de sobreajuste, aún cuando el AUC de validación muestra una mejora desde 0.6458 hasta alcanzar un máximo de 0.9864 en la época 13. Se experimentan caídas abruptas en momentos clave, como el descenso a 0.8380 en la época 4 y a 0.9087 en la época 26 que delatan una inconsistencia en la capacidad discriminatoria del modelo con datos no vistos.

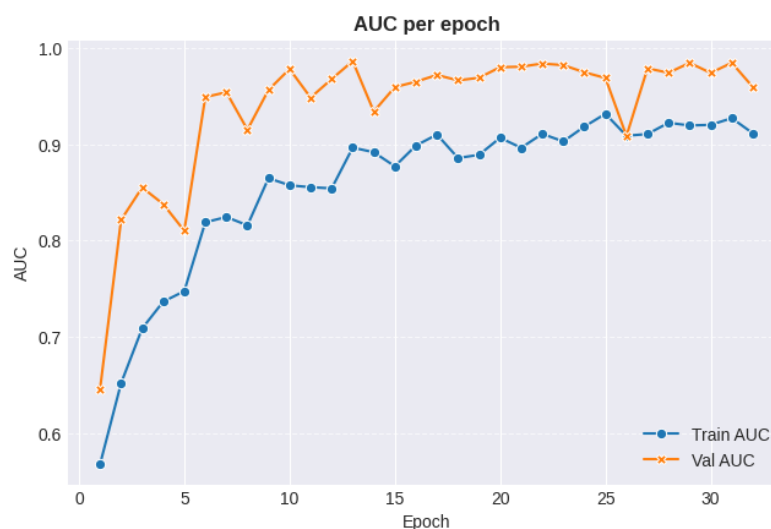


Figura 6.18. Evolución del AUC del modelo UNet

De forma similar, la precisión de validación con una trayectoria ascendente que culmina en 94.25 % en la época 22 presenta oscilaciones alarmantes, como la caída de 81.64 % a 75.49 % entre las épocas 4 y 5, y

de 93.96 % a 87.55 % entre las épocas 13 y 14, donde el modelo presenta momentáneos deterioros en su predicción, creando brechas variables con las métricas de entrenamiento que exponen una generalización irregular.

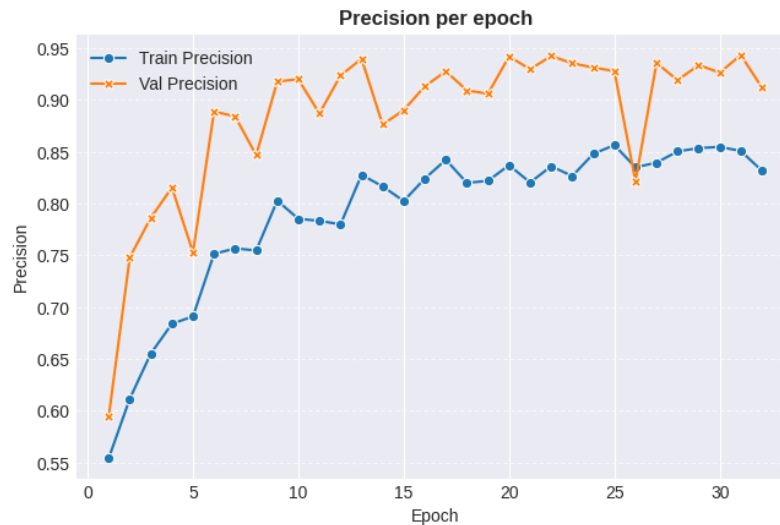


Figura 6.19. Evolución de la precisión del modelo UNet

El análisis combinado de la métrica *recall* y el tiempo de ejecución (ms/step) indica un comportamiento de aprendizaje con inconsistencias temporales y de generalización, donde el *recall* de validación demuestra una tendencia ascendente general desde un 57.90 % inicial hasta un máximo de 94.23 % en la época 22, evidenciando la capacidad del modelo para detectar casos positivos. No obstante, fluctuaciones significativas como la caída de 81.58 % a 75.76 % entre las épocas 4 y 5, y el descenso de 93.88 % a 87.18 % entre las épocas 13 y 14 relevan problemas intermitentes de sensibilidad y sobreajustes momentáneos a los datos de entrenamiento.



Figura 6.20. Evolución del *recall* del modelo UNet

Del mismo modo, el tiempo de ejecución por paso experimenta una alta variabilidad, iniciando con 546 ms en la primera época hasta 750 ms en la época 10, con valores que oscilan entre 323 ms y 578 ms en épocas

posteriores, reflejando una eficiencia computacional irregular que sugiere procesos en segundo plano relacionados con operaciones de entrada y salida, *checkpointing* y ajustes en el *scheduler* de aprendizaje, donde los picos de tiempo no correlacionan directamente con mejoras en el rendimiento del modelo, indicando que la inestabilidad en los tiempos de ejecución podrían contribuir a las irregularidades de aprendizaje al crear condiciones inconsistentes para la convergencia del gradiente durante el entrenamiento.

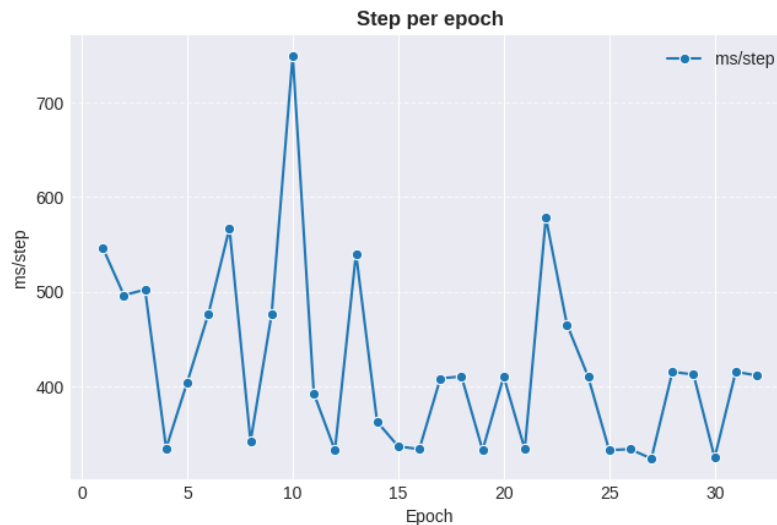


Figura 6.21. Evolución del *step* por época del modelo UNet

Por otra parte, se combina la arquitectura DenseNet121 para segmentación de imágenes con un decodificador en forma de U como muestra la figura 6.13. La parte principal del modelo utiliza una versión preentrenada con los primeros 100 niveles de procesamiento bloqueados para evitar modificaciones, extrayendo características relevantes en cinco escalas diferentes. La primera escala se representa con un tamaño de 128x128 y 64 filtros, la segunda de 64x64 con 128 filtros, la tercera 32x32 con 256 filtros, la cuarta 16x16 con 512 filtros y la última de 8x8 con 1024 filtros. La parte del decodificador del modelo aumenta progresivamente el tamaño de la imagen mediante operaciones que combinan aumento de resolución usando filtros de 2x2 de paso 2 con información de las capas correspondientes de la parte analizadora, aplicando posteriormente filtros de 3x3 con relleno uniforme y función de activación ReLU junto con normalización de datos, utilizando sucesivamente 512, 256, 128 y 64 filtros en cada etapa.

La fase final emplea operaciones orientadas al aumento de resolución con 32 filtros seguidas de filtrado (32 filtros) para recuperar el tamaño original de la imagen, culminando con una capa final que usa filtros 1x1 y 2 filtros con activación softmax para generar la máscara de segmentación con 2 canales. El modelo DenseNet121 tiene un total de 19,209,666 parámetros, de los cuales 18,319,362 parámetros son entrenables y 890,304 no entrenables principalmente por la normalización por lotes.

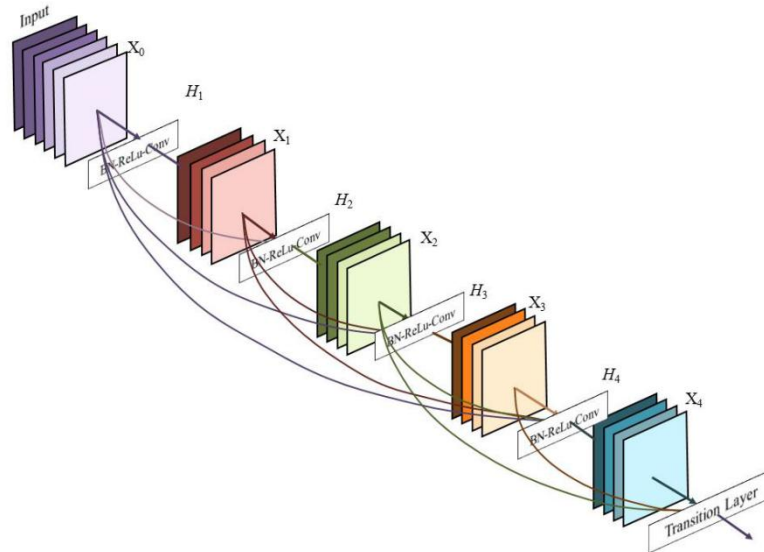


Figura 6.22. Arquitectura DenseNet121. Tomada de [49]

El modelo presenta una evolución inicial positiva con una rápida disminución de la pérdida de validación desde 0.10265 en la época 1 hasta su mínimo de 0.03082 en la época 13 (ver figura 6.23), acompañada de una mejora en la exactitud de validación que alcanzó su máximo del 96.14 % (ver figura 6.24). Sin embargo, a partir de la época 5 se evidencia un sobreajuste, donde la pérdida de entrenamiento continúa disminuyendo de 0.1259 a 0.0690 pero la pérdida de validación comienza a fluctuar y aumentar, lo cual indica que el modelo está memorizando los datos de entrenamiento en vez de generalizar patrones.

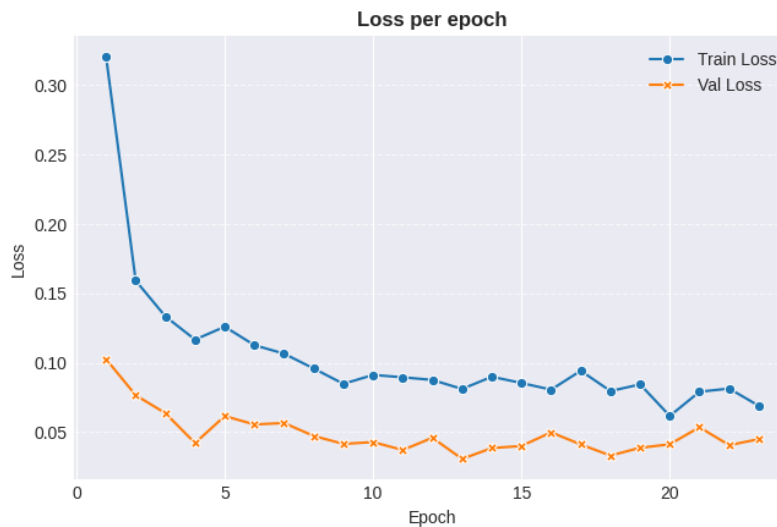


Figura 6.23. Evolución de la pérdida del modelo DenseNet121

Esta divergencia progresiva entre la pérdida baja de entrenamiento y la pérdida de validación estancada, y luego en aumento culmina con la activación del *early stopping* en la época 23, restaurando los pesos de la época 13 que representan el mejor equilibrio entre aprendizaje y generalización del modelo. Aunque el modelo demostró una rápida convergencia, completando el entrenamiento en 27.01 minutos, este rendimiento se vio afectado por la presencia de sobreajuste. La falta de generalización sugiere que la

rapidez del entrenamiento comprometió su efectividad, llevando al modelo a aprender correlaciones irrelevantes presentes en los datos de entrenamiento.

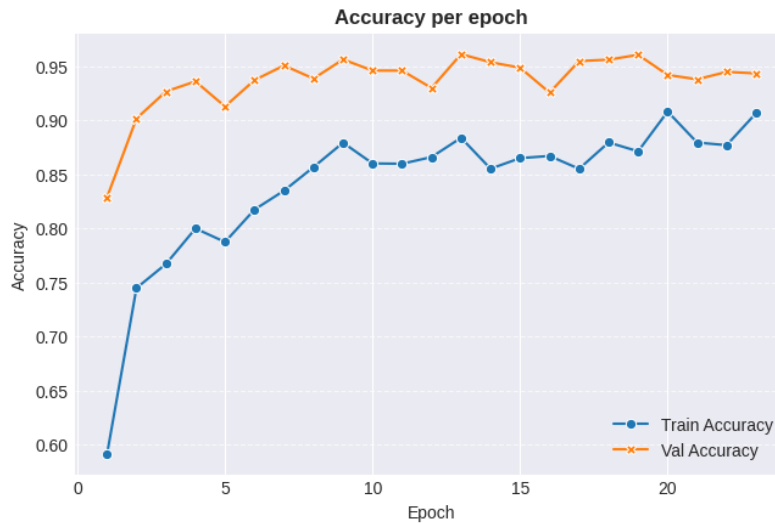


Figura 6.24. Evolución de la exactitud del modelo DenseNet121

Como se presenta en la figura 6.25, el modelo demuestra un desempeño alto en cuanto a la métrica AUC, alcanzando un valor máximo de validación de 0.9900 en la época 13, lo que indica la capacidad de distinguir entre clases positivas y negativas.

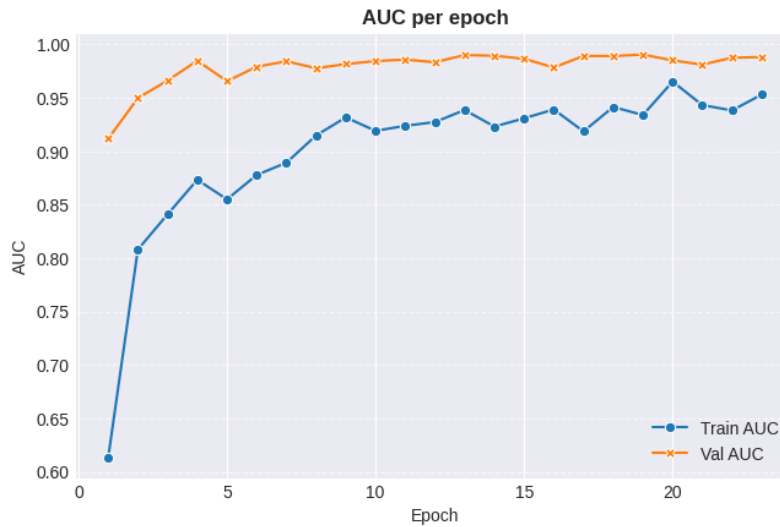


Figura 6.25. Evolución del AUC del modelo DenseNet121

Esta capacidad de discriminación se respalda con una alta precisión de validación que llegó al 96.14 % en la misma época, confirmando que cuando el modelo predice una clase positiva tiene alta probabilidad de ser correcta. Sin embargo, se observa que, aunque el AUC de validación se mantuvo consistente alto, por encima de 0.98 en la mayoría de las épocas, la precisión experimentó gran parte de volatilidad en la fase

final del entrenamiento, fluctuando entre 92.58 % y 96.08 %, indicando que mientras la capacidad general de clasificación AUC se mantuvo robusta, la confiabilidad de las predicciones positivas en precisión mostró mayor sensibilidad de sobreajuste.

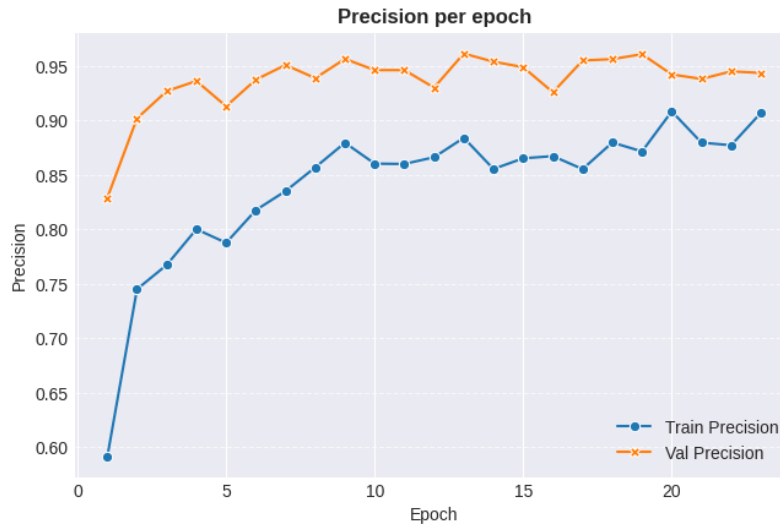


Figura 6.26. Evolución de la precisión del modelo DenseNet121

A pesar de que el modelo exhibió un rendimiento alto en la métrica de *recall*, alcanzando un valor máximo de validación del 96.14 % en la época 13, la aparición de un claro sobreajuste se manifiesta con fluctuaciones posteriores en esta métrica. El *recall* de validación mostró una notable volatilidad, cayendo hasta 92.58 % en la época 16 y oscilando inconsistentemente entre 93.81 % y 96.08 % en lugar de mantener una mejora estable, indicando que el modelo comenzó a memorizar ruido y patrones específicos del conjunto de entrenamiento.

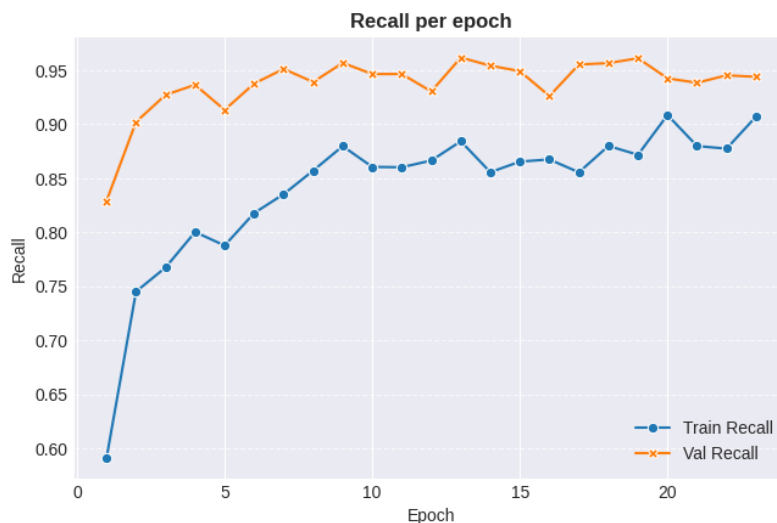


Figura 6.27. Evolución del *recall* del modelo DenseNet121

En adición, esta inestabilidad en el *recall* correlaciona con los picos en el tiempo en ms/step, como los 459 ms en la época 11 y 410 ms en la época 23, donde las operaciones de parada temprana y reducción

de la tasa de aprendizaje interrumpieron la eficiencia del entrenamiento, demostrando que se llegó a un punto de rendimientos decrecientes donde el entrenamiento adicional no mejoró la generalización, sino que aumentó la volatilidad del modelo, comprometiendo su consistencia predictiva. la evolución de las métricas *step* por época del modelo DenseNet121 se presenta en la figura 6.28.

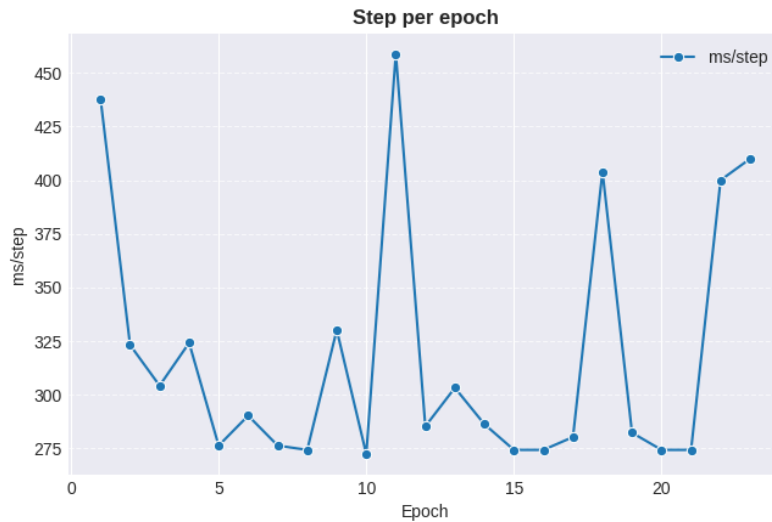


Figura 6.28. Evolución del *step* por época del modelo DenseNet121

7. DESEMPEÑO DE LOS MODELOS DE SEGMENTACIÓN DE IMÁGENES HISTOPATOLÓGICAS

Una vez finalizado el entrenamiento, se evaluó el rendimiento de los modelos de segmentación UNet + ResNet 50, ResNet50, UNet y DenseNet121 utilizando el conjunto de prueba de imágenes histopatológicas H&E. El modelo UNet + ResNet50 demostró un rendimiento robusto en la segmentación semántica binaria, logrando una exactitud del 94.98 % junto con un Dice y F1 Score del 94.64 %, que refleja una superposición casi perfecta con las máscaras de referencia, manteniendo un equilibrio entre ser preciso y exhaustivo, sin sacrificar una métrica por la otra. La métrica *recall* de 94.02 % garantiza una mínima tasa de falsos negativos, crucial en el diagnóstico médico, mientras que su precisión con 95.28 % asegura confiabilidad al reducir falsas alarmas. La capacidad discriminativa del modelo queda confirmada por su curva ROC, que alcanzó un valor AUC de 0.9876, indicando una probabilidad del 98.76 % de distinguir correctamente entre clases positivas y negativas. La matriz de confusión confirma este equilibrio, con solo un 4.16 % de falsos positivos y un 5.98 % de falsos negativos, consolidando al modelo como una herramienta robusta y prometedora para asistir la evaluación clínica. En la figura 7.1 se muestra en (a) la matriz de confusión porcentual y en (b) la curva de ROC con el valor de AUC del modelo UNet + ResNet50.

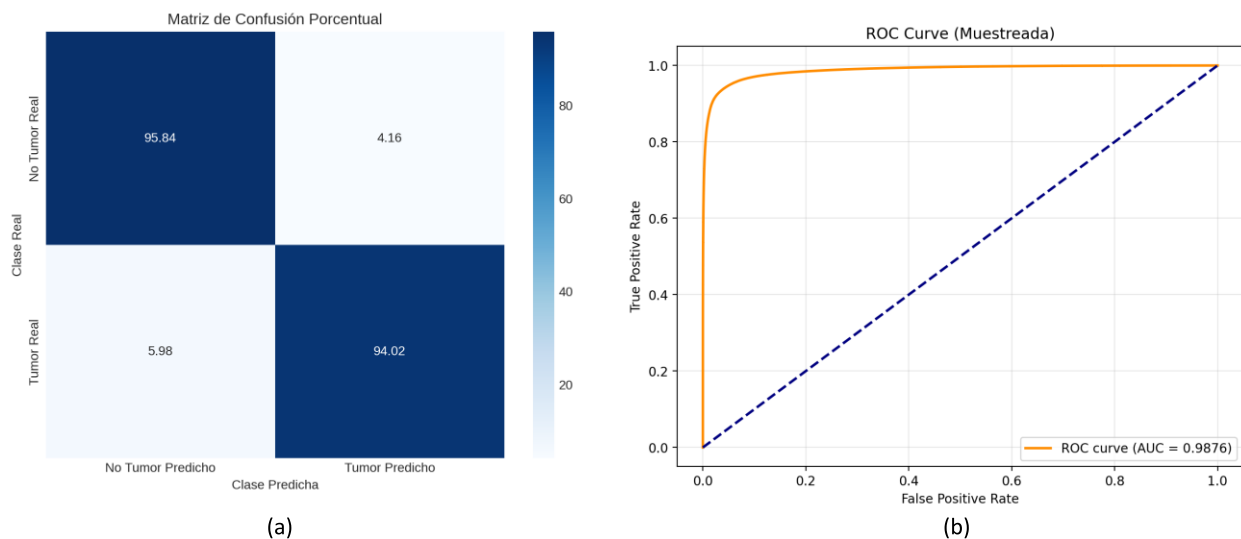


Figura 7.1. Desempeño general del modelo UNet + ResNet50

El modelo ResNet50 presenta un rendimiento global alto con exactitud de 94.13 %, Dice score de 94.30 %, F1-score de 93.91 % y capacidad para identificar la totalidad del ganglio con un *recall* de 95.94 %. A pesar de su capacidad discriminativa, evidenciada por una curva ROC con un AUC de 0.9830, el modelo es afectado por un sobreajuste claro, lo que se manifiesta en una precisión de 91.96 % relativamente menor causada por un crítico porcentaje de falsos positivos de 7.48 %. Esto indica que el modelo alucina estructuras y genera ruido o produce manchas fuera del área real del ganglio, debido a que memoriza patrones específicos del conjunto de entrenamiento en vez de aprender características generales de los ganglios linfáticos. La métrica de AUC confirma que el modelo separa parcialmente las clases, pero no mitiga su tendencia a generar falsas alarmas, relevando que su alto rendimiento en clasificación no se traduce directamente en una segmentación espacialmente precisa. La figura 7.2 muestra en (a) la matriz de confusión porcentual y en (b) la curva de ROC con el valor de AUC del modelo ResNet50.

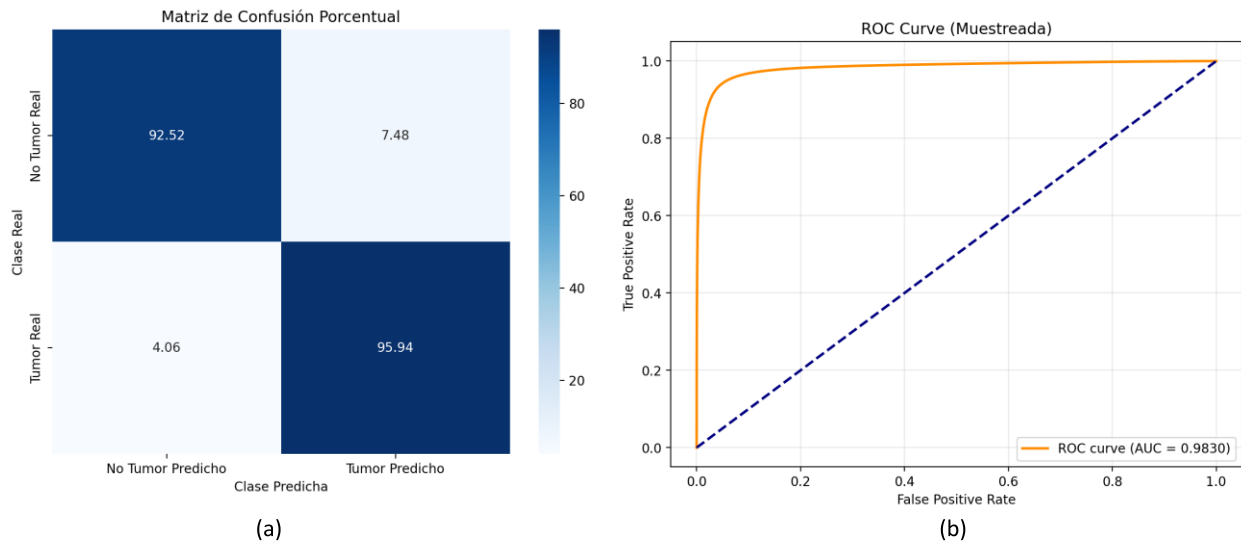


Figura 7.2. Desempeño general del modelo ResNet50

A pesar de que las métricas del modelo UNet son numéricamente altas, con una exactitud del 92.96 %, un *recall* del 93.46 % que muestra alta sensibilidad para detectar pixeles de ganglios, una precisión del 91.76 % y un Dice Score de 92.60 % que indica una superposición cercana con las máscaras de referencia, la curva ROC, que alcanzó un AUC de 0.9772, releva una capacidad de discriminación alta. Sin embargo, esta métrica por sí sola no contradice la evidencia de sobreajuste. La presencia de este problema implica que el rendimiento global es engañoso para la tarea de segmentación médica. Cuando el modelo se enfrente a nuevas imágenes que no ha visto antes, no generalizará correctamente y, al haber memorizado los datos de entrenamiento, su rendimiento caerá drásticamente, lo que traducirá en un aumento de los falsos negativos más allá de 6.54 %, arriesgado a omitir ganglios cancerosos, y de los falsos positivos superando el 7.49 %, creando ganglios irreales, lo cual compromete su utilidad clínica a pesar de la prometedora métrica de AUC. La figura 7.3 presenta: (a) la matriz de confusión porcentual y (b) la curva ROC del modelo.

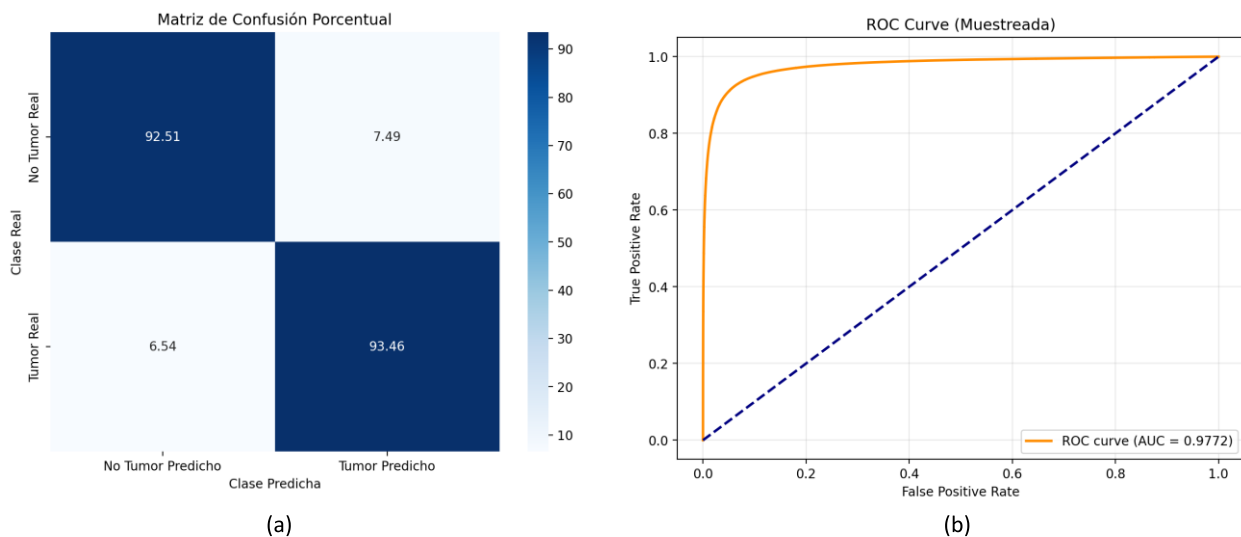


Figura 7.3. Desempeño general del modelo UNet

Por otra parte, las métricas del modelo DenseNet121 son relativamente altas, reportando una exactitud del 95.23 %, sensibilidad del 94.97 % y precisión del 94.91 %, respaldadas por la matriz de confusión porcentual de la figura 7.4 (a), su curva ROC con un AUC de 0.9838, presentada en la figura 7.4 (b), parece confirmar su desempeño alto desempeño, pero en realidad esta métrica constituye la evidencia más aparente de un severo sobreajuste. Estos resultados indican que el modelo ha memorizado los datos de entrenamiento en lugar de aprender patrones generalizables, probablemente debido a un conjunto de datos insuficiente para su arquitectura compleja. La capacidad de discriminación reflejada en el AUC resulta ser una visión que no se mantendrá en entornos clínicos. En consecuencia, aunque la puntuación F1 y Dice actual es de 94.94 %, se anticipa una drástica caída en estas métricas al evaluar el modelo con otros datos no vistos, comprometiendo así su utilidad clínica cuando deba enfrentarse a la variabilidad natural de las imágenes médicas.

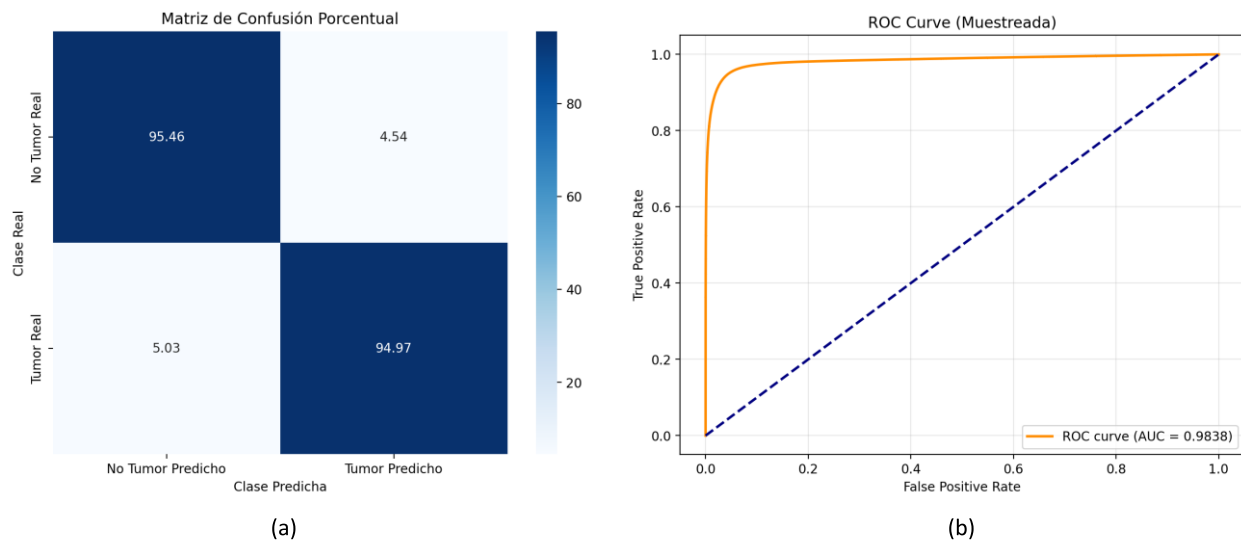


Figura 7.4. Desempeño general del modelo DenseNet121

Con el fin de ilustrar de manera representativa el desempeño de los modelos de segmentación en la tarea de detección de metástasis, se ha seleccionado una muestra de cinco imágenes histológicas (H&E) y representativas provenientes del conjunto de pruebas. Esta elección se justifica por la necesidad de realizar un análisis comparativo detallado y contrastar la precisión de las diferentes arquitecturas frente a la máscara de verdad, facilitando la identificación de patrones de aciertos y errores. La figura 7.5 presenta un análisis comparativo del rendimiento de diversos modelos de segmentación para la detección de metástasis en ganglios linfáticos con cáncer de mama. En la sección (a) se muestran cinco parches de prueba, seguidos de la máscara verdadera de segmentación en (b). Las predicciones de los distintos modelos se despliegan en orden: el modelo híbrido propuesto (UNet + ResNet50) en (c), el modelo ResNet50 en (d), el modelo UNet en (e) y el modelo DenseNet121 en (f). El análisis global demuestra que la arquitectura híbrida propuesta fue la que exhibió una mayor robustez, superando a los modelos individuales. Las segmentaciones por este modelo presentan la menor intensidad de error a nivel de píxel en comparación con las máscaras de referencia, donde los píxeles en rojo (valor de error igual a 1.0) indican una discrepancia total, es decir, casos en los que el modelo realizó una predicción incorrecta con respecto al valor verdadero. Esta mínima presencia de píxeles rojos en los resultados del modelo híbrido confirma visualmente su precisión superior y efectividad para esta tarea diagnóstica.

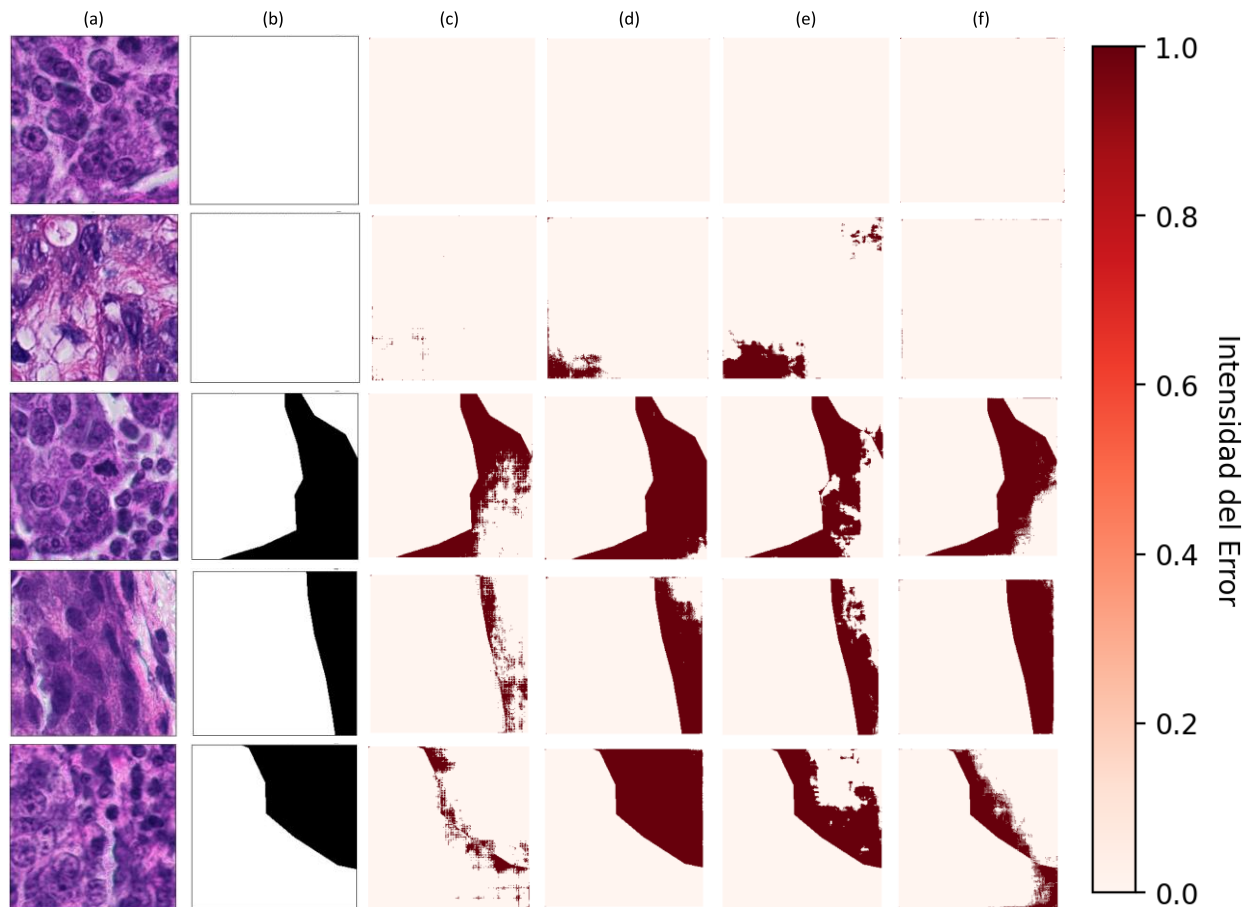


Figura 7.5. Análisis comparativo del rendimiento de los modelos de segmentación

La tabla 5 cuantifica el desempeño de los distintos modelos de segmentación comparados, reportando los valores obtenidos en las métricas de Jaccard, precisión, *recall*, Y F1 score para los cinco parches de prueba analizados. Para la interpretación de los resultados, se emplea la siguiente nomenclatura: el modelo híbrido propuesto (UNet + ResNet50) se identifica como MP, mientras que los modelos de referencia se denotan como M1 (ResNet50), M2 (UNet) y M3 (DenseNet121). El análisis de los resultados demuestra que el modelo MP establece un referente en la segmentación de ganglios linfáticos en imágenes de H&E, superando de manera consistente y significativa a los modelos M1, M2 y M3 en todas las métricas clave. Su superioridad en la métrica de Jaccard, el indicador directo de la precisión de la segmentación, donde MP logra los valores más altos en cuatro de los cinco parches, con un rendimiento perfecto (100 %) en el parche 1 y una robustez en entornos desafiantes con un 75.88 % en el parche 3, superando por más de 10 puntos porcentuales a M1. Esta ventaja se mantiene en el F1-score, que sintetiza el equilibrio entre precisión y *recall*, confirmando que MP no solo es preciso al identificar píxeles correctos con una precisión consistente entre 92 % y 100 %, sino también completo al no pasar por alto regiones de interés, con *recall* constantemente por encima del 95 %, alcanzando el 99.96 % en el parche 3.

El análisis de las medias consolida esta superioridad. MP obtiene la puntuación global más alta en Jaccard (91.86 %), precisión (92.73 %) y F1-score (95.53 %), lo que refleja su robustez y equilibrio general superiores. Aunque en la media de *recall* (99.03 %), M1 lo supera ligeramente (99.29 %), el desempeño integral de MP, capturado por las métricas clave de Jaccard, precisión y F1 sigue siendo incomparable.

Mientras los otros modelos muestran vulnerabilidades y fluctuaciones notables en su desempeño, particularmente M1 en los parches 2 y 5, la solidez y fiabilidad del modelo híbrido MP lo posicionan como la solución para esta tarea de diagnóstico asistido. Por otra parte, el análisis de la desviación estándar (SD) revela que el modelo MP es el más consistente y robusto. Su fortaleza principal radica en presentar la menor variabilidad en la mayoría de las métricas clave. En concreto, para las métricas de Jaccard (9.81 %), Precisión (9.92 %) y F1-score (5.59 %), MP registra la desviación estándar más baja de todos los métodos. Esto significa que su rendimiento es notablemente estable y predecible a través de los diferentes parches, sin sufrir caídas bruscas. En la métrica *Recall*, aunque M1 presenta una SD ligeramente menor (1.55 % frente a 1.87 % de MP), MP mantiene una variabilidad muy baja y, crucialmente, la media de *Recall* más alta (99.03 %), lo que confirma que su alta capacidad de detección es a la vez potente y confiable. La importancia de esta consistencia se hace especialmente evidente en el F1-score, la medida armónica que balancea Precisión y *Recall*. La baja SD de MP (5.59 %) indica que su desempeño general combinado es el más equilibrado y fiable. En contraste, métodos como M1 o M3, a pesar de poder alcanzar puntuaciones medias altas en algunos casos, muestran desviaciones estándar considerablemente mayores, por ejemplo, 16.62 % para Jaccard en M1 o 13.26 % en M3. Esta mayor SD señala una inconsistencia y su rendimiento puede ser excelente en un escenario y significativamente peor en otro, lo que los hace menos predecibles.

Tabla 5. Comparación de métricas de segmentación

	Jaccard (%)				Precisión (%)				Recall (%)				F1-score (%)			
	MP	M1	M2	M3	MP	M1	M2	M3	MP	M1	M2	M3	MP	M1	M2	M3
Parche 1	100	100	100	99.94	100	100	100	100	100	100	100	99.94	100	100	100	99.97
Parche 2	99.76	96.52	90.72	99.86	100	100	100	100	99.76	96.52	90.72	99.86	99.88	98.23	95.13	99.93
Parche 3	75.88	65.33	73.41	70.71	75.90	65.34	74.18	70.78	99.96	99.98	98.61	99.85	86.28	79.03	84.67	82.84
Parche 4	92.57	80.65	83.88	77.66	92.80	80.67	83.89	77.67	99.73	99.97	100	99.97	96.14	89.29	91.24	87.42
Parche 5	91.07	64.91	75.66	82.31	94.97	64.92	75.66	86.05	95.69	99.99	100	94.99	95.33	78.72	86.15	90.30
Media	91.86	81.48	84.73	86.10	92.73	82.19	86.75	86.90	99.03	99.29	97.87	98.92	95.53	89.05	91.44	92.09
SD	9.81	16.62	10.96	13.26	9.92	17.46	12.65	13.12	1.87	1.55	4.04	2.20	5.59	10.14	6.34	7.65

El modelo híbrido propuesto (MP) demuestra una precisión superior en la estimación de carga tumoral, al exhibir los errores de predicción más bajos y consistentes en comparación con los modelos M1, M2 y M3, tal como se presenta en la tabla 6. Mientras los otros modelos cometen errores drásticamente altos, superando el 20 % e incluso el 35 %, como M1 en el parche 5, MP se mantiene notablemente cercano al valor real, con un error mínimo del 0 % en el parche 1, un 0.50 % en el crítico parche 5, y un máximo de 5.78 % en el parche 4. Este desempeño contrasta con la de M1, M2 y M3, cuyos errores son frecuentemente de magnitud mayores, lo que confirma que el modelo híbrido MP no solo es más preciso, sino también el más robusto, evitando graves sobrestimaciones o subestimaciones que caracterizan a sus alternativas, posicionándolo como una herramienta para la cuantificación tumoral en entornos de diagnóstico.

Tabla 6. Estimación de la carga tumoral

	Valor real (%)	MP (%)	Error MP (%)	M1 (%)	Error M1 (%)	M2 (%)	Error M2 (%)	M3 (%)	Error M3 (%)
Parche 1	100	100	0	100	0	100	0	99.94	0.06
Parche 2	100	99.76	0.24	96.52	3.48	90.72	9.28	99.86	0.14
Parche 3	64.78	85.33	20.54	99.14	34.36	86.11	21.33	91.39	26.61
Parche 4	77.39	83.16	5.78	95.90	18.51	92.25	14.86	99.60	22.22
Parche 5	64.82	65.32	0.50	99.84	35.02	85.67	20.85	71.56	6.74

Para continuar con la prueba del modelo propuesto UNet + ResNet50 para la segmentación de ganglios linfáticos, se compara ahora con trabajos del estado del arte donde se hayan involucrado estructuras basadas en aprendizaje profundo y se comparó el desempeño de las métricas, como se muestra en la tabla 7. La selección de 10 trabajos se justifica porque incluye modelos arquitectónicos clave publicados en revistas de alto impacto Q1 y Q2, lo que garantiza que la comparación se realice contra los referentes más citados y avanzados en el área, ofreciendo una base de comparación sólida y significativa para validar el desempeño del modelo híbrido propuesto [50].

El modelo híbrido propuesto (UNet + ResNet50) se constituye como una solución robusta y avanzada para la segmentación de ganglios linfáticos en imágenes histopatológicas H&E del dataset CAMELYON, demostrando una superioridad integral y un equilibrio en todas las métricas de evaluación. Al alcanzar un *accuracy* del 94.98 %, no solo supera en más de un punto porcentual a DSNet (93.51 %), sino que establece una brecha mayor con otros modelos Q1 como ConcatNet (84.1 %) e HI-Net (88.4 %). Sin embargo, su verdadera capacidad radica en la combinación sinérgica de un *recall* del 94.02 % y una precisión del 95.28 %, lo que se traduce en un puntaje F1 y Dice del 94.64 %. Esta métrica compuesta es fundamental, ya que confirma que el modelo es eficaz para identificar la mayor cantidad posible de ganglios verdaderos (alta sensibilidad) y para minimizar drásticamente las detecciones falsas (alta precisión), un desafío crucial en el análisis médico. Esta ventaja es abrumadora cuando se compara con otros modelos Dice, como CGS-Net (70.04 %) y DeepLabV3+MobileNet (75 %), a los que supera por más de 19 y 24 puntos porcentuales respectivamente. Finalmente, su potencial de discriminación queda validada por un AUC de 98.76 %, la más alta reportada, lo que indica una probabilidad mínima de confusión entre clases. En conjunto, estos resultados validan que la arquitectura híbrida, al combinar la capacidad de segmentación precisa de UNet con el desempeño de extracción de características profundas de ResNet50, ofrece alto rendimiento, estableciendo un nuevo estado del arte para esta tarea de diagnóstico exigente.

Tabla 7. Comparación de rendimiento entre el modelo propuesto y trabajos del estado del arte publicados en revistas de alto impacto Scopus

Trabajo	Scopus	Técnica	Accuracy	Recall	Precision	F1-score	Dice score	AUC
Modelo híbrido propuesto	-	UNet+ResNet50	94.98 %	94.02 %	95.28 %	94.64 %	94.64 %	98.76 %
[51]	Q1	CA-SegNet	-	-	83.95 %	-	81.49 %	-
[52]	Q1	ConcatNet	84.1 %	82.0 %	-	-	-	92.4 %

[53]	Q1	DSNet	93.51 %	-	-	-	-	-
[54]	Q1	DeepLabV3 + MobileNet V3	-	-	-	-	75 %	-
[55]	Q1	CGS-Net MIT-B1	-	-	-	-	70.04 % ± 12.28 %	98.06 % ± 0,02 %
[56]	Q1	HI-Net	88.4 %	-	-	-	-	-
[57]	Q2	FCN modificada	-	83.8%	89.6 %	84.4 %	-	-
[58]	Q1	Deeplabv3+ + ResNet V1	91.16 %	62.55 %	-	-	-	-
[59]	Q1	DeeplabV3	-	-	-	-	-	96.9 %
[60]	Q1	DCNN	-	-	-	-	-	95 %

8. CONCLUSIONES

La etapa del análisis del conjunto de imágenes de histología fue crucial y se centró en la preparación del dataset CAMELYON16 para el desarrollo de los modelos de segmentación. Debido a las grandes dimensiones de las imágenes de lámina completa (WSI), se implementó una estrategia de extracción de parches de 256x256 píxeles que, combinada con un método automatizado de filtrado multietapa, permitió descartar áreas irrelevantes, artefactos o fondos grises homogéneos. El análisis exploratorio de datos reveló una alta heterogeneidad intermuestral en la distribución del tejido tumoral, con porcentajes que variaron desde un mínimo de 3.9 % hasta un predominio absoluto del 98.4 % en algunas muestras. Para mitigar las variaciones cromáticas causadas por diferencias en la tinción, se aplicó el método de normalización de color propuesto por Macenko, transformando las imágenes al espacio de densidad óptica y asegurando una representación visualmente consistente para que las características extraídas fueran robustas. Finalmente, se implementó un proceso de aumento de datos, incluyendo transformaciones geométricas y ajustes de color en el conjunto de entrenamiento a las 26,450 muestras para incrementar la diversidad y mejorar la capacidad de generalización de los modelos de aprendizaje profundo frente a las condiciones heterogéneas encontradas en entornos reales.

El análisis comparativo en la etapa de entrenamiento, realizado bajo las mismas condiciones experimentales y configuraciones para garantizar la imparcialidad, establece que el modelo híbrido UNet + ResNet50 superó consistentemente a las arquitecturas ResNet50, UNet y DenseNet121 en términos de robustez y capacidad de generalización. Mientras que el modelo híbrido demostró un aprendizaje efectivo y una mejora consistente con un buen equilibrio entre sesgo y varianza, los demás modelos mostraron signos significativos de sobreajuste que comprometieron su generalización. Específicamente, ResNet50 (entrenado en 26.82 minutos) y DenseNet121 (entrenado en 27.01 minutos) fueron ligeramente más rápidos que el modelo híbrido (31.98 minutos), pero sufrieron de divergencia entre las curvas de entrenamiento y validación, llevando a la activación de la parada temprana. UNet, por su parte, fue el modelo más lento (45.08 minutos) y mostró inestabilidad pronunciada, con fluctuaciones notables en sus métricas y una capacidad inconsistente de generalización. En contraste con la inestabilidad de los otros modelos, la arquitectura híbrida alcanzó las métricas predictivas más altas, logrando un AUC de validación 0.9961 y máximos de precisión y *recall* de 0.9708, superando los valores máximos de DenseNet121 (AUC 0.9884, precisión y *recall* 96.14 %), ResNet50 (AUC 0.9884, precisión 95.18 % y *recall* 95.08 %) y UNet (AUC 0.9864, precisión 94.25 % y *recall* 94.23 %). Además, el modelo híbrido destacó por ser la arquitectura con menor cantidad de parámetros (15,635,234) en comparación con ResNet50 (43,980,306), UNet (31,055,362) y DenseNet121 (19,209,666), logrando un equilibrio superior entre potencia de representación y eficiencia computacional sin sacrificar la capacidad de generalización.

El modelo híbrido UNet + ResNet50 se establece como la solución más robusta y equilibrada para la segmentación de ganglios linfáticos, superando consistentemente a sus alternativas individuales (ResNet50, UNet y DenseNet121) y a modelos avanzados del estado del arte. Si bien DenseNet121 reportó una exactitud numéricamente superior (95.23 %) y un Dice score (94.94 %), esta métrica, junto con su AUC de 0.9838, constituye la evidencia más aparente de un sobreajuste, lo que anticipa una drástica caída de rendimiento en datos no vistos, comprometiendo su utilidad clínica. De manera similar, ResNet50, aunque con un alto *recall* (95.94 %), presenta una precisión menor del 91.96 % causada por un crítico porcentaje de falsos positivos (7.48 %), indicando que el modelo alucina estructuras. UNet también exhibió vulnerabilidades al sobreajuste (AUC de 0.9772), lo que podría llevar a un aumento de falsos negativos (omitiendo ganglios cancerosos) y falsos positivos en la práctica clínica. En contraste, el modelo híbrido

propuesto mantiene una superioridad integral en las métricas clave, logrando un Dice score de 94.64 % y una exactitud de 94.98 % sin estos problemas de generalización. Su desempeño no solo es más preciso y menos propenso a errores de segmentación a nivel de píxel que ResNet50, UNet y DenseNet121, sino que también minimiza los errores de predicción de carga tumoral, manteniéndose con un error máximo de 5.78 %, mientras que los otros modelos cometen errores drásticamente altos, superando el 20 % e incluso el 35 %. Además, el modelo híbrido supera a referentes del estado de arte, como DSNet (93.51 % de exactitud) y CGS-Net (70.04 % de Dice score), consolidando su AUC del 98.76 % como el valor más alto reportado, lo que valida su potencial de discriminación y establece un nuevo referente en la segmentación de ganglios linfáticos.

La innovación de esta arquitectura UNet + ResNet50 no está en su estructura base, sino en su sistema de regularización adaptativa y optimización específica. Se implementó una estrategia diferenciada que incluye la congelación selectiva de capas, *dropout* variable según la dimensión de los mapas de características, y tasas de *dropout* espacial decrecientes en las conexiones directas. Esta configuración, junto al uso de convoluciones dilatadas en el puente y una distribución intencional de parámetros (55 % en el codificador), aborda problemas clave como el sobreajuste y la pérdida de información. La innovación y contribución metodológica reside en un marco de ajuste fino y regularización heterogénea que maximiza su rendimiento en tareas de segmentación especializada, ofreciendo un paradigma replicable para mejorar arquitecturas híbridas existentes.

9. TRABAJOS FUTUROS

A futuro, se propone profundizar en las siguientes líneas de investigación:

- Integrar el modelo híbrido desarrollado en una herramienta de diagnóstico asistido por computador, orientada a la segmentación de ganglios linfáticos y a la estimación de la carga tumoral en imágenes de cáncer de mama teñidas con H&E. Esta implementación proporciona un apoyo al especialista, no solo mediante la identificación de las estructuras ganglionares, sino también de la cuantificación objetiva de la afectación metastásica.
- Validación de cohortes multicéntricas más grandes y diversas para evaluar la robustez y generalización del modelo.
- Desarrollar y probar técnicas de aumento de datos y normalización para manejar de forma eficaz las variaciones entre scanners de diapositivas y tinciones histológicas
- Explorar arquitecturas de Deep learning más avanzadas o especializadas para mejorar la precisión en la delimitación de los bordes de las metástasis.
- Implementar un modelo de estimación de carga tumoral cuantitativa que integre automáticamente los resultados de la segmentación (volumen y número de metástasis segmentadas).
- Desarrollar un método para la segmentación y clasificación celular individual dentro de conglomerados tumorales superpuestos utilizando variantes del algoritmo Watershed.
- Integrar el modelo de segmentación con sistemas de pronóstico clínico establecidos, como el Residual Cancer Burden (RCB), para predecir la respuesta a la quimioterapia neoadyuvante.
- Crear un modelo predictivo simplificado que combine los hallazgos de la segmentación (como el porcentaje de área metastásica) con otros factores pronósticos clave (el estado de los receptores hormonales) para estimar la supervivencia.
- Validar y aplicar el modelo de segmentación para detectar y cuantificar metástasis en otros órganos comunes.
- Investigar la correlación entre las características cuantitativas extraídas de las metástasis segmentadas (como la forma, textura o la densidad celular) y la agresividad del tumor o la respuesta al tratamiento.

10. REFERENCIAS BIBLIOGRÁFICAS

- [1] C. Kaushal, S. Bhat, D. Koundal, and A. Singla, "Recent Trends in Computer Assisted Diagnosis (CAD) System for Breast Cancer Diagnosis Using Histopathological Images," *IRBM*, vol. 40, no. 4, pp. 211–227, Aug. 2019, doi: 10.1016/j.irbm.2019.06.001.
- [2] S. yang Tang, L. Li, Y. lin Li, A. yuan Liu, M. jun Yu, and Y. ping Wan, "Distribution and location of Daxx in cervical epithelial cells with high risk human papillomavirus positive," *Diagn Pathol*, vol. 9, no. 1, Jan. 2014, doi: 10.1186/1746-1596-9-1.
- [3] F. Li, J. Ma, T. Wen, Z. Tian, and H. N. Liang, "HI-Net: A novel histopathologic image segmentation model for metastatic breast cancer via lightweight dataset construction," *Heliyon*, vol. 10, no. 19, p. e38410, Oct. 2024, doi: 10.1016/J.HELIYON.2024.E38410.
- [4] Y. Lu *et al.*, "Prediction of breast cancer metastasis by deep learning pathology," *IET Image Process*, vol. 17, no. 2, pp. 533–543, Feb. 2023, doi: 10.1049/IPR2.12652.
- [5] R. Rashmi, K. Prasad, C. B. K. Udupa, and V. Shwetha, "A Comparative Evaluation of Texture Features for Semantic Segmentation of Breast Histopathological Images," *IEEE Access*, vol. 8, pp. 64331–64346, 2020, doi: 10.1109/ACCESS.2020.2984522.
- [6] Y. Wang *et al.*, "Improved breast cancer histological grading using deep learning," *Annals of Oncology*, vol. 33, no. 1, pp. 89–98, Jan. 2022, doi: 10.1016/j.annonc.2021.09.007.
- [7] C. V. Lakshmi Priya, V. G. Biju, B. R. Vinod, and S. Ramachandran, "Deep learning approaches for breast cancer detection in histopathology images: A review," *Cancer Biomark*, vol. 40, no. 1, pp. 1–25, May 2024, doi: 10.3233/CBM-230251.
- [8] B. M. Priego-Torres, D. Sanchez-Morillo, M. A. Fernandez-Granero, and M. Garcia-Rojo, "Automatic segmentation of whole-slide H&E stained breast histopathology images using a deep convolutional neural network architecture," *Expert Syst Appl*, vol. 151, p. 113387, Aug. 2020, doi: 10.1016/J.ESWA.2020.113387.
- [9] Y. Feng *et al.*, "Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis," *Genes Dis*, vol. 5, no. 2, pp. 77–106, Jun. 2018, doi: 10.1016/J.GENDIS.2018.05.001.
- [10] A. Juhong *et al.*, "Super-resolution and segmentation deep learning for breast cancer histopathology image analysis," *Biomedical Optics Express*, Vol. 14, Issue 1, pp. 18-36, vol. 14, no. 1, pp. 18–36, Jan. 2023, doi: 10.1364/BOE.463839.
- [11] R. Rashmi, K. Prasad, and C. B. K. Udupa, "Breast histopathological image analysis using image processing techniques for diagnostic purposes: A methodological review," *Journal of Medical Systems 2021 46:1*, vol. 46, no. 1, pp. 1–24, Dec. 2021, doi: 10.1007/S10916-021-01786-9.
- [12] S. I. Kim, C. R. Cassella, and K. T. Byrne, "Tumor Burden and Immunotherapy: Impact on Immune Infiltration and Therapeutic Outcomes," *Front Immunol*, vol. 11, p. 629722, Feb. 2021, doi: 10.3389/FIMMU.2020.629722/BIBTEX.
- [13] R. Krithiga and P. Geetha, "Breast Cancer Detection, Segmentation and Classification on Histopathology Images Analysis: A Systematic Review," *Archives of Computational Methods in Engineering*, vol. 28, no. 4, pp. 2607–2619, Jun. 2021, doi: 10.1007/S11831-020-09470-W/TABLES/2.
- [14] L. Jose, S. Liu, C. Russo, A. Nadort, and A. Di Ieva, "Generative Adversarial Networks in Digital Pathology and Histopathological Image Processing: A Review," *J Pathol Inform*, vol. 12, no. 1, p. 43, Jan. 2021, doi: 10.4103/JPI.JPI_103_20.
- [15] K. P. Trayes and S. E. H. Cokenakes, "Breast Cancer Treatment," *Am Fam Physician*, vol. 104, no. 2, pp. 171–178, Aug. 2021, Accessed: Nov. 10, 2024. [Online]. Available:

- <https://www.aafp.org/pubs/afp/issues/2021/0800/p171.html>
- [16] A. F. Faulana *et al.*, “Review: Aplikasi Histopatologi Untuk Praktik Forensik,” *Borneo Journal of Medical Laboratory Technology*, vol. 6, no. 2, pp. 529–539, Apr. 2024, doi: 10.33084/BJMLT.V6I2.6582.
- [17] C. L. Srinidhi, O. Ciga, and A. L. Martel, “Deep neural network models for computational histopathology: A survey,” *Med Image Anal*, vol. 67, p. 101813, Jan. 2021, doi: 10.1016/J.MEDIA.2020.101813.
- [18] M. Z. Hoque, A. Keskinarkaus, P. Nyberg, and T. Seppänen, “Stain normalization methods for histopathology image analysis: A comprehensive review and experimental comparison,” *Information Fusion*, vol. 102, p. 101997, Feb. 2024, doi: 10.1016/J.INFFUS.2023.101997.
- [19] M. Muñoz-Aguirre, V. F. Ntasis, S. Rojas, and R. Guigó, “PyHIST: A Histological Image Segmentation Tool,” *PLoS Comput Biol*, vol. 16, no. 10, p. e1008349, Oct. 2020, doi: 10.1371/JOURNAL.PCBI.1008349.
- [20] A. Mahbod *et al.*, “CryoNuSeg: A dataset for nuclei instance segmentation of cryosectioned H&E-stained histological images,” *Comput Biol Med*, vol. 132, p. 104349, May 2021, doi: 10.1016/J.COMPBIOMED.2021.104349.
- [21] X. Ling *et al.*, “Towards a Comprehensive Benchmark for Pathological Lymph Node Metastasis in Breast Cancer Sections,” Nov. 2024, doi: 10.57760/sciencedb.16442.
- [22] T. A. Azevedo Tosta, P. R. de Faria, L. A. Neves, and M. Z. do Nascimento, “Computational normalization of H&E-stained histological images: Progress, challenges and future potential,” *Artif Intell Med*, vol. 95, pp. 118–132, Apr. 2019, doi: 10.1016/J.ARTMED.2018.10.004.
- [23] J. Boschman *et al.*, “The utility of color normalization for AI-based diagnosis of hematoxylin and eosin-stained pathology images,” *Journal of Pathology*, vol. 256, no. 1, pp. 15–24, Jan. 2022, doi: 10.1002/PATH.5797.
- [24] A. Mumuni and F. Mumuni, “Data augmentation: A comprehensive survey of modern approaches,” *Array*, vol. 16, p. 100258, Dec. 2022, doi: 10.1016/J.ARRAY.2022.100258.
- [25] T. Kumar, R. Brennan, A. Mileo, and M. Bendeche, “Image Data Augmentation Approaches: A Comprehensive Survey and Future Directions,” *IEEE Access*, vol. 12, pp. 187536–187571, 2024, doi: 10.1109/ACCESS.2024.3470122.
- [26] M. Alenezi, M. Akour, A. Hussien, and M. Z. Al-Saad, “Test suite effectiveness: An indicator for open source software quality,” *2016 2nd International Conference on Open Source Software Computing, OSSCOM 2016*, Feb. 2017, doi: 10.1109/OSSCOM.2016.7863677.
- [27] M. S. Yim *et al.*, “Deep learning-driven macroscopic AI segmentation model for brain tumor detection via digital pathology: Foundations for terahertz imaging-based AI diagnostics,” *Heliyon*, vol. 10, no. 22, Nov. 2024, doi: 10.1016/j.heliyon.2024.e40452.
- [28] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 770–778, Dec. 2015, doi: 10.1109/CVPR.2016.90.
- [29] J. Jiang, X. Chen, G. Tian, and Y. Liu, “ViG-UNet: Vision Graph Neural Networks for Medical Image Segmentation,” *Proceedings - International Symposium on Biomedical Imaging*, vol. 2023-April, Jun. 2023, doi: 10.1109/ISBI53787.2023.10230496.
- [30] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 2261–2269, Aug. 2016, doi: 10.1109/CVPR.2017.243.
- [31] W. R. Drioua, N. Benamrane, and L. Sais, “Breast Cancer Histopathological Images Segmentation Using Deep Learning,” *Sensors 2023, Vol. 23, Page 7318*, vol. 23, no. 17, p. 7318, Aug. 2023, doi:

- 10.3390/S23177318.
- [32] Q. Liu *et al.*, “Multi-module UNet++ for colon cancer histopathological image segmentation,” *Sci Rep*, vol. 15, no. 1, pp. 1–19, Dec. 2025, doi: 10.1038/S41598-025-13636-6;SUBJMETA.
- [33] Z. Lu *et al.*, “BreasTDLUSeg: A coarse-to-fine framework for segmentation of breast terminal duct lobular units on histopathological whole-slide images,” *Computerized Medical Imaging and Graphics*, vol. 118, p. 102432, Dec. 2024, doi: 10.1016/J.COMPAMEDIMAG.2024.102432.
- [34] A. Pedersen *et al.*, “H2G-Net: A multi-resolution refinement approach for segmentation of breast cancer region in gigapixel histopathological images,” *Front Med (Lausanne)*, vol. 9, p. 971873, Sep. 2022, doi: 10.3389/FMED.2022.971873/BIBTEX.
- [35] Y. Varnava, K. Jakate, R. Garnett, D. Androustos, P. N. Tyrrell, and A. Khademi, “Out-of-distribution generalization for segmentation of lymph node metastasis in breast cancer,” *Scientific Reports 2024 15:1*, vol. 15, no. 1, pp. 1–16, Jan. 2025, doi: 10.1038/s41598-024-80495-y.
- [36] K. Ikromjanov *et al.*, “Region Segmentation of Whole-Slide Images for Analyzing Histological Differentiation of Prostate Adenocarcinoma Using Ensemble EfficientNetB2 U-Net with Transfer Learning Mechanism,” *Cancers (Basel)*, vol. 15, no. 3, p. 762, Feb. 2023, doi: 10.3390/CANCERS15030762.
- [37] T. Zhao, C. Fu, M. Tie, C. W. Sham, and H. Ma, “RGSB-UNet: Hybrid Deep Learning Framework for Tumour Segmentation in Digital Pathology Images,” *Bioengineering 2023, Vol. 10, Page 957*, vol. 10, no. 8, p. 957, Aug. 2023, doi: 10.3390/BIOENGINEERING10080957.
- [38] R. Hernández Sampieri, C. Feránadez Collado, and M. D. P. Baptista Lucio, “Metodología de la investigación,” *Metodología de la investigación*, p. 91, 2014, Accessed: Oct. 01, 2025. [Online]. Available: <https://dialnet.unirioja.es/servlet/libro?codigo=775008&info=resumen&idioma=SPA>
- [39] G. Shmueli and O. R. Koppius, “Predictive Analytics in Information Systems Research1,” *Management Information Systems Quarterly*, vol. 35, no. 3, pp. 553–572, Sep. 2011, doi: 10.2307/23042796.
- [40] J. G. Yu *et al.*, “Prototypical multiple instance learning for predicting lymph node metastasis of breast cancer from whole-slide pathological images,” *Med Image Anal*, vol. 85, p. 102748, Apr. 2023, doi: 10.1016/J.MEDIA.2023.102748.
- [41] S. Heidari Gheshlaghi, M. Aryal, N. Yahya Soltani, and M. Ganji, “Explainability-based graph augmentation for out-of-distribution robustness in digital pathology,” *Knowl Based Syst*, vol. 320, p. 113640, Jun. 2025, doi: 10.1016/J.KNOSYS.2025.113640.
- [42] E. Bütün, M. Uçan, and M. Kaya, “Automatic detection of cancer metastasis in lymph node using deep learning,” *Biomed Signal Process Control*, vol. 82, p. 104564, Apr. 2023, doi: 10.1016/J.BSPC.2022.104564.
- [43] M. Veta, P. J. van Diest, R. Kornegoor, A. Huisman, M. A. Viergever, and J. P. W. Pluim, “Automatic Nuclei Segmentation in H&E Stained Breast Cancer Histopathology Images,” *PLoS One*, vol. 8, no. 7, p. e70221, Jul. 2013, doi: 10.1371/JOURNAL.PONE.0070221.
- [44] M. Macenko *et al.*, “A method for normalizing histology slides for quantitative analysis,” *Proceedings - 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2009*, pp. 1107–1110, 2009, doi: 10.1109/ISBI.2009.5193250.
- [45] M. Isik, “Comprehensive empirical evaluation of feature extractors in computer vision,” *PeerJ Comput Sci*, vol. 10, p. e2415, Nov. 2024, doi: 10.7717/PEERJ-CS.2415/TABLE-6.
- [46] W. C. Yee *et al.*, “Performance Analysis of Color Normalization Methods in Histopathology Images,” *2022 IEEE International Conference on Automatic Control and Intelligent Systems, I2CACIS 2022 - Proceedings*, pp. 147–151, 2022, doi: 10.1109/I2CACIS54679.2022.9815475.
- [47] H. M. Hama, T. Sh. Abdulsamad, and S. M. Omer, “Houseplant leaf classification system based on

- deep learning algorithms,” *Journal of Electrical Systems and Information Technology* 2024 11:1, vol. 11, no. 1, pp. 1–15, Apr. 2024, doi: 10.1186/S43067-024-00141-5.
- [48] N. Mohamadiazar, A. Ebrahimian, and H. Hosseiny, “Integrating deep learning, satellite image processing, and spatial-temporal analysis for urban flood prediction,” *J Hydrol (Amst)*, vol. 639, p. 131508, Aug. 2024, doi: 10.1016/J.JHYDROL.2024.131508.
- [49] B. Wang *et al.*, “An Efficient Cloud Classification Method Based on a Densely Connected Hybrid Convolutional Network for FY-4A,” *Remote Sensing* 2023, Vol. 15, Page 2673, vol. 15, no. 10, p. 2673, May 2023, doi: 10.3390/RS15102673.
- [50] S. Ríos, S. N. S., F. A., G. A., and J. R. M., “Bibliografía,” *Trabajos de Estadística* 1953 4:3, vol. 4, no. 3, pp. 393–397, Oct. 1953, doi: 10.1007/BF03013502.
- [51] F. He, W. Wang, L. Ren, Y. Zhao, Z. Liu, and Y. Zhu, “CA-SegNet: A channel-attention encoder–decoder network for histopathological image segmentation,” *Biomed Signal Process Control*, vol. 96, p. 106590, Oct. 2024, doi: 10.1016/J.BSPC.2024.106590.
- [52] Y. W. Jin, S. Jia, A. B. Ashraf, and P. Hu, “Integrative Data Augmentation with U-Net Segmentation Masks Improves Detection of Lymph Node Metastases in Breast Cancer Patients,” *Cancers (Basel)*, vol. 12, no. 10, pp. 1–13, Oct. 2020, doi: 10.3390/CANCERS12102934.
- [53] L. Wang, T. Song, T. Katayama, X. Jiang, T. Shimamoto, and J. S. Leu, “Deep regional metastases segmentation for patient-level lymph node status classification,” *IEEE Access*, vol. 9, pp. 129293–129302, 2021, doi: 10.1109/ACCESS.2021.3113036.
- [54] Y. Varnava, K. Jakate, R. Garnett, D. Androutsos, P. N. Tyrrell, and A. Khademi, “Out-of-distribution generalization for segmentation of lymph node metastasis in breast cancer,” *Sci Rep*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/S41598-024-80495-Y.
- [55] J. Juybari, J. Hamilton, C. Chen, A. Khalil, and Y. Zhu, “Context-guided segmentation for histopathologic cancer segmentation,” *Sci Rep*, vol. 15, no. 1, pp. 1–12, Dec. 2025, doi: 10.1038/S41598-025-86428-7;SUBJMETA.
- [56] F. Li, J. Ma, T. Wen, Z. Tian, and H. N. Liang, “HI-Net: A novel histopathologic image segmentation model for metastatic breast cancer via lightweight dataset construction,” *Heliyon*, vol. 10, no. 19, p. e38410, Oct. 2024, doi: 10.1016/j.heliyon.2024.e38410.
- [57] M. A. Khalil, Y. C. Lee, H. C. Lien, Y. M. Jeng, and C. W. Wang, “Fast Segmentation of Metastatic Foci in H&E Whole-Slide Images for Breast Cancer Diagnosis,” *Diagnostics (Basel)*, vol. 12, no. 4, Apr. 2022, doi: 10.3390/DIAGNOSTICS12040990.
- [58] B. M. Priego-Torres, D. Sanchez-Morillo, M. A. Fernandez-Granero, and M. Garcia-Rojo, “Automatic segmentation of whole-slide H&E stained breast histopathology images using a deep convolutional neural network architecture,” *Expert Syst Appl*, vol. 151, p. 113387, Aug. 2020, doi: 10.1016/J.ESWA.2020.113387.
- [59] J. Li, A. Osseyran, R. Hekster, S. Rudinac, V. Codreanu, and D. Podareanu, “Improving the speed and quality of cancer segmentation using lower resolution pathology images,” *Multimed Tools Appl*, vol. 83, no. 4, pp. 11999–12015, Jan. 2024, doi: 10.1007/S11042-023-15984-9/FIGURES/6.
- [60] Z. Guo *et al.*, “A Fast and Refined Cancer Regions Segmentation Framework in Whole-slide Breast Pathological Images,” *Sci Rep*, vol. 9, no. 1, pp. 1–10, Dec. 2019, doi: 10.1038/S41598-018-37492-9;TECHMETA.