UNIVERSITAT POLITÈCNICA DE VALÈNCIA

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Industrial

Diseño y desarrollo de un sistema de segmentación de la epidermis en imágenes histopatológicas de melanoma Spitzoide mediante redes encoder-decoder residuales

Trabajo Fin de Máster

Máster Universitario en Ingeniería Biomédica

AUTOR/A: Viault , Pauline Florence Marie

Tutor/a: Naranjo Ornedo, Valeriana

Cotutor/a: Colomer Granero, Adrián

Cotutor/a externo: GOLFE SAN MARTIN, ALEJANDRO

CURSO ACADÉMICO: 2021/2022

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

ESCUELA TÉCNICA SUPERIOR INGENIERÍA INDUSTRIAL VALENCIA

TRABAJO FIN DE MASTER EN INGENIERÍA BIOMÉDICA

DISEÑO Y DESARROLLO DE UN SISTEMA DE SEGMENTACIÓN DE LA EPIDERMIS EN IMÁGENES HISTOPATOLÓGICAS DE MELANOMA SPITZOIDE MEDIANTE REDES ENCODER-DECODER RESIDUALES

AUTORA: PAULINE VIAULT

TUTORA: VALERY NARANJO ORNEDO

COTUTORES: ADRIAN COLOMER GRANERO
ALEJANDRO GOLFE SAN MARTIN

Curso Académico: 2021-22

Agradecimientos

Para la realización de este trabajo de fin de master en ingeniería biomédica primero quiero dar las gracias a mis tutores Valery Naranjo Ornedo y Alejandro Golfe San Martin con una atención especial a Alejandro que me ha acompañado durante todas mis prácticas, resolviendo mis dudas y dándome muchos consejos útiles e ideas novedosas. Siempre me han guiado para alcanzar resultados de precisión y robustos en un área donde tienen unos conocimientos grandes y una experiencia amplia. Agradezco también a Rocío del Amor por todo su apoyo al principio de mis prácticas en el laboratorio.

También me gustaría mostrar mi sincero agradecimiento a Adrián Colomer Granero que me ha guiado mediante reuniones del equipo de histopatología y seguido en todo este proceso. Agradezco a Cristian por su apoyo importante con el manejo de los recursos informáticos del laboratorio.

Doy las gracias también al patólogo Andrés David Mosquera Zamudio por su ayuda y sus explicaciones histológicas y anatómicas de la epidermis y su supervisión de la anotación de las imágenes para la base de datos. La involucración de personal médico como el permite la elaboración de proyectos como este trabajo de fin de master.

Al equipo pedagógico y todos los profesores del máster de ingeniería biomédica de la Universidad politécnica de Valencia que me han proporcionado los conocimientos y habilidades necesarias para llevar a cabo esta tarea y empezar en el dominio de la investigación científica.

A todo el laboratorio CVB Lab por el uso de sus infraestructuras, su base de datos de imágenes y su equipo dinámico y a la vanguardia en algoritmos de inteligencia artificial.

Resumen

La segmentación de la epidermis en imágenes histopatológicas con tinción hematoxilina eosina es una tarea esencial en el análisis que realizan los patólogos para el diagnóstico de tumores melanocíticos de Spitz. Sin embargo, analizar a mano cada muestra puede resultar una tarea tediosa para el especialista a la par que subjetiva. Por ello, la introducción de la patología digital puede suponer un ahorro de tiempo así como una mayor robustez y precisión en la segmentación de grandes estructuras de biopsias como, por ejemplo, la epidermis. Esto permite que el patólogo pueda centrarse en el análisis en profundidad del tejido y de sus células. Además, la segmentación automática de la epidermis provee datos numéricos (cuantitativos) que no es posible obtener en la patología clásica.

En este trabajo el alumno trabajará un enfoque basado en métodos de *Deep Learning* para detectar la capa de epidermis en biopsias de piel. Se diseñará una arquitectura de red neuronal convolucional encoder-decoder para resolver el problema de segmentación. Se explorará la creación de un modelo predictivo a dos niveles de resolución y la adición de capas residuales para lograr una segmentación automática precisa. La exitosa consecución del presente TFM brindará la posibilidad de poder desarrollar sistemas de detección de tumores melanocíticos de Spitz mediante, entre otros, la cuantificación automática de la epidermis.

Palabras clave: Segmentación de la epidermis; Melanoma de Spitz; *Deep Learning*; *U-Net*; *Residual U-Net*; Patología digital; Redes neuronales convolucionales; *Dice* score

Abstract

Epidermis segmentation in histopathological images with haematoxylin eosin staining is an essential task in the pathologist's analysis for the diagnosis of Spitz melanocytic tumours. However, analysing each sample by hand can be a tedious and subjective task for the specialist. Therefore, the introduction of digital pathology can lead to time savings as well as increased robustness and accuracy in the segmentation of large biopsy structures such as the epidermis. This allows the pathologist to focus on the in-depth analysis of the tissue and its cells. Moreover, the automatic segmentation of the epidermis provides numerical (quantitative) data that is not possible to obtain in classical pathology.

In this work the student will work on an approach based on Deep Learning methods to detect the epidermis layer in skin biopsies. An encoder-decoder convolutional neural network architecture will be designed to solve the segmentation problem. The creation of a predictive model at two levels of resolution and the addition of residual layers will be explored to achieve accurate automatic segmentation. The successful completion of this TFM will provide the possibility to develop Spitz melanocytic tumour detection systems using, among others, automatic quantification of the epidermis.

Keywords: Epidermis segmentation; Spitz melanocytic tumour; Deep Learning; U-Net; Residual U-Net; Digital pathology; Convolutional Neural Networks; Dice score

índice de contenidos

Índice de ilustraciones

Índice de tablas

I. Memoria

1. Introducción

1.1. Introducción general

Hoy en día los canceres de la piel y especialmente el melanoma son cada vez más frecuentes [1]. Actualmente, cada año se producen en el mundo entre 2 y 3 millones de cánceres de piel no melanoma y 132.000 cánceres de piel melanoma [2]. Uno de cada tres cánceres diagnosticados es un cáncer de piel. La disminución de los niveles de ozono atmosféricos debida a la actividad humana contaminante hace que llegue más radiación UV solar y eso provocará 300.000 casos adicionales de cáncer de piel no melanoma y 4.500 de melanoma. Aunque se presenta más habitualmente en adultos mayores, es el tercer tipo de cáncer detectado en adolescentes y adultos jóvenes (de 15 a 39 años) [3].

El melanoma maligno, aunque es mucho menos frecuente que los cánceres de piel no melanoma, es la principal causa de muerte por cáncer de piel. El diagnóstico del melanoma se realiza mediante las biopsias de piel que inspeccionadas al microscopio por el patólogo permiten observar las células cancerosas y el desarrollo del tumor. Entre las lesiones cutáneas, las melanocíticas de Spitz representan un reto para el diagnóstico del patólogo [4]. Las lesiones melanocíticas son tumores que se desarrollan a partir de los melanocitos. Los diagnósticos erróneos del melanoma de Spitz pueden tener graves repercusiones para los pacientes.

Con la ayuda de los escáneres digitales de histología, las biopsias de las muestras de tejido se pueden digitalizar para crear la *whole slide image* (WSI en este trabajo). Sin embargo, la tarea del patólogo para el diagnóstico de cáncer de piel es tediosa. Además, el diagnóstico manual es subjetivo y a menudo da lugar a una variabilidad en muestras etiquetadas por un mismo observador y entre distintos observadores [5]. Se intenta proporcionar nuevas herramientas digitales y automáticas en un proceso llamado patología digital para proporcionar resultados objetivos, fiables y reproducibles. Son resultados que permiten añadir un análisis cuantitativo al análisis cualitativo realizado por los médicos. Una muestra de tejido cutáneo presenta tres partes: la epidermis, la dermis y la hipodermis [6], como se ve en la Ilustración 1. En la mayoría de los casos de melanoma, la clasificación del cáncer puede hacerse a través de las características morfológicas de las células atípicas en la epidermis. Por lo tanto, la segmentación de la epidermis es un paso importante antes de realizar otros análisis.

El presente trabajo tiene como finalidad implementar distintas arquitecturas de redes neuronales artificiales para la segmentación automática de la epidermis en imágenes de biopsias de piel.

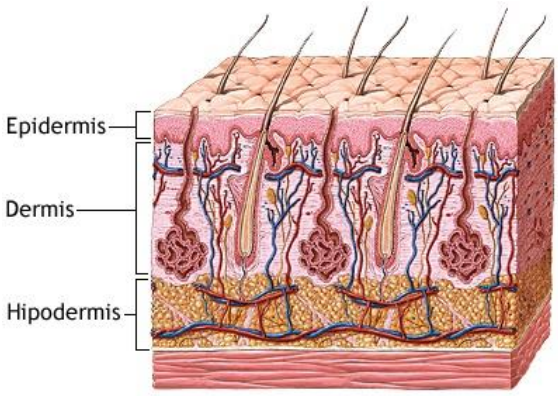


Ilustración 1: Esquema de las capas de la piel [7]

1.2. Contexto del TFM

En este apartado se presentan los conceptos que permiten arrojar luz sobre el contexto y los motivos de este trabajo. Este estudio se inscribe en la implementación de la patología digital usando imágenes de biopsias de piel. Es de gran interés en el mundo científico y el proyecto europeo CLARIFY justifica también el trabajo presentado aquí.

1.2.1. Introducción a la patología digital

Como previamente comentado, la tarea del patólogo para confirmar o establecer un diagnóstico en casos de enfermedades complejas, como puede ser la oncología, es tediosa. De hecho, tienen que observar muchas muestras a ojo y fijarse en las estructuras de interés. El tejido u órgano de interés llega a un laboratorio de histopatología que realiza las etapas de fijación, inclusión, microtomía y tinción, como se ve en la Ilustración 2. Estas etapas sirven para lo siguiente:

- La **fijación** impide la putrefacción y conserva la composición molecular de la muestra.
- La **inclusión** proporciona consistencia para facilitar el corte en secciones finas. Se sustituye el agua (componente mayoritario de células y tejidos) por parafina.
- La **microtomía** es una técnica que permite obtener secciones histológicas suficientemente finas para que la luz del microscopio las atraviese y puedan visualizarse.
- Las células sin teñir tienen un 70% de agua y en su mayoría son transparentes. La **tinción** es una técnica auxiliar utilizada en microscopía para mejorar el contraste en la imagen microscópica.

ILUSTRACIÓN 2: PASOS DE OBTENCIÓN DE MUESTRAS BIOLÓGICAS [8]

La tinción más utilizada en diagnóstico clínico es la de hematoxilina-eosina (denominada H&E en este trabajo). Es una mezcla de hematoxilina que tiñe los núcleos de azul y eosina que tiñe los citoplasmas de rosa. Permite visualizar la anatomía de la mayoría de los tejidos analizados. Existen otras tinciones de referencias para motivos más específicos como la tinción tricromática de Masson para el tejido conectivo (colágeno, tejido fibroso, cartílago) o el azul Prusia de Perl para marcar el hierro.

La muestra estudiada se obtiene mediante biopsia cutánea. La biopsia cutánea es un examen médico que consiste en extirpar un trozo fino y pequeño de la piel del paciente para analizarlo en un laboratorio de anatomía patológica gracias a su observación con el microscopio. Se puede realizar de varias formas según la región en la que se encuentra la lesión de interés. Los diferentes tipos de biopsia cutánea más utilizados son la biopsia por afeitado, la biopsia por punción y la biopsia tras una escisión.

Con las nuevas técnicas de procesado de imágenes y los algoritmos de inteligencia artificial se trata de proveer a los patólogos de herramientas digitales para facilitar su trabajo. Estos algoritmos pueden ser muy precisos, reducir el tiempo que el patólogo debe dedicar a cada muestra y también pueden servir para la formación de nuevos patólogos. Estos algoritmos suponen una revolución en la patología actual. Las perspectivas de mejora facilitando el diagnóstico en este sector son muy potentes.

Las herramientas automatizadas de análisis de imágenes permiten también realizar un análisis cuantitativo que suele ser complicado de llevar a cabo a mano cuando el patólogo observa las imágenes con el microscopio. Para automatizar el análisis patológico se empieza a emplear inteligencia artificial en los laboratorios como descrito en el artículo [9]. En la sección que sigue se describe con mayor detalle la estructura de la piel humana para comprender las especificidades de la epidermis.

1.2.2. Estructura de la piel humana

La piel es el órgano más extendido del cuerpo humano. Tiene muchas funciones esenciales al organismo como constituir una barrera de protección contra el entorno, servir como órgano sensorial y jugar un papel muy importante en la termorregulación.

Está compuesta por tres capas:

- La **epidermis** como capa exterior de la piel. Se compone de cuatro capas principales: la capa córnea, la capa granular, la capa espinosa y la capa basal como se ve en la Ilustración 4. La epidermis no tiene vasos sanguíneos y depende de la dermis para nutrirse.
- La **dermis**, la capa intermedia, se encuentra debajo de la epidermis. Es el principal tejido de soporte de la piel y es responsable de su solidez. Es el lugar donde se encuentran los folículos pilosos, las glándulas sudoríparas y las glándulas sebáceas. Todos estos componentes aseguran la hidratación y la nutrición de la piel y participan en la protección del organismo contra las agresiones.
- La **hipodermis** es la capa más profunda de la piel. Está formada por células grasas, los adipocitos. Almacena energía y proporciona aislamiento para conservar el calor corporal.

Dentro de las principales células de la piel, se pueden citar los queratinocitos que producen la queratina y citocinas, los melanocitos que producen la melanina, la sustancia que da su color a la piel y las células dendríticas de Langerhans que interaccionan con el sistema inmune. La Ilustración 3 permite visualizar una esquematización de estas células dentro de la epidermis, se ve que los queratinocitos son el tipo celular mayoritario que la compone.

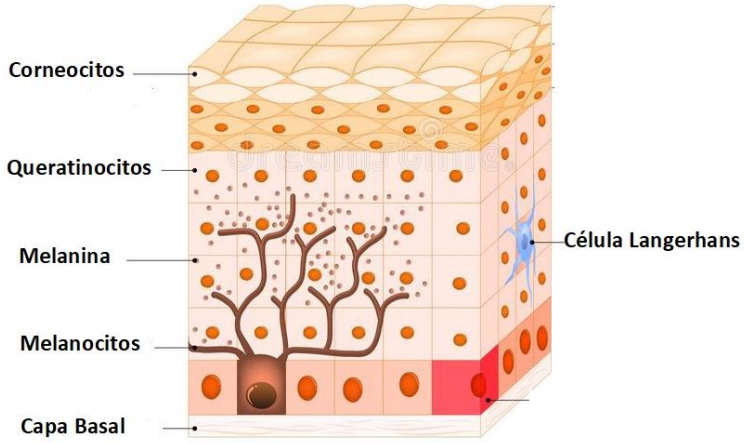


ILUSTRACIÓN 3: ESQUEMA DE LAS CÉLULAS DE LA EPIDERMIS [10]

En la Ilustración 4 y la Ilustración 5 con más detalle, se ven las capas que componen la epidermis. La epidermis es un epitelio escamoso estratificado que contiene de cuatro a cinco capas dependiendo de su ubicación. Presenta células que están muy pegadas las unas a las otras para proveer su papel protector a la piel. En más detalle las diferentes capas son:

- El *stratum basal* o capa de células basales, es la capa más profunda y cercana a la dermis. Contiene melanocitos y una única fila de queratinocitos de forma cuboides. Los queratinocitos evolucionan y maduran desde esta capa hacia el exterior para crear las siguientes capas arriba. A medida que evolucionan van cambiando de forma entre las diferentes capas.
- El *stratum spinosum* o capa de células espinosas, es la mayor parte de la epidermis con varias capas de células fuertemente unidas entre sí.
- El *stratum granulosum* o capa de células granulares, es una capa con células aplanadas que comienzan a perder sus núcleos.
- El *stratum lucidum* es una capa únicamente presente en las plantas de los pies y las palmas de las manos y formada de células inmortalizadas.
- El *stratum corneum* o capa de queratina, es la capa exterior que sirve de recubrimiento protector. Está compuesta de células muertas que se pueden desprender. Esta capa permite regular la pérdida de agua al impedir la evaporación interna de fluidos mediante su contenido de lípidos.

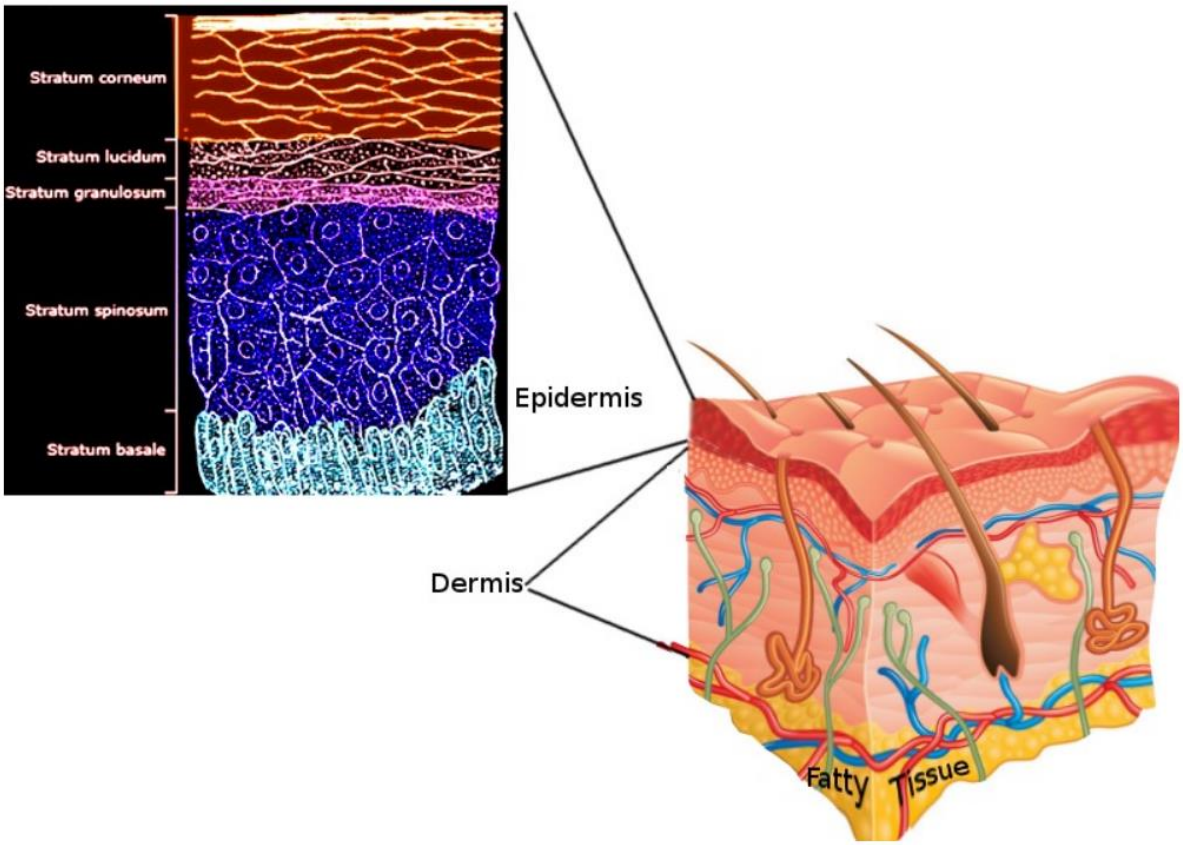


ILUSTRACIÓN 4: ESQUEMA DE LA ESTRUCTURA DE LA EPIDERMIS [11]

ILUSTRACIÓN 5: DIFERENTES CAPAS DE LA EPIDERMIS EN IMAGEN HISTOLÓGICA DE PIEL DE RATÓN [12]

La piel tiene apéndices de la epidermis en la dermis que complica la localización exacta del límite dermoepidérmico. La zona de unión de la epidermis y de la dermis presenta estructuras llamadas papilas dérmicas y crestas epidérmicas como representado en la Ilustración 6. Una papila dérmica es una estructura pequeña digitiforme de la dermis que entra en la epidermis. Las crestas epidérmicas se encuentran presentes mayoritariamente en la piel de las manos y los pies. Los vasos sanguíneos de las papilas dérmicas son importantes para nutrir las estructuras de la epidermis que carece de vasos sanguíneos.

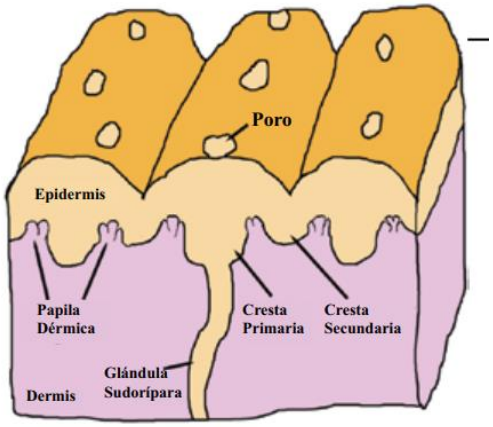


ILUSTRACIÓN 6: ESQUEMAS DE LAS CRESTAS EPIDÉRMICAS Y DE LAS PAPILAS DÉRMICAS [13]

Como indicado previamente, los folículos pilosos pertenecen mayoritariamente a la dermis pero cierta parte llamada el infundíbulo es un elemento de tejido epitelial. El infundíbulo es el tercio superior del folículo. El tercio medio es el istmo y recibe las glándulas sebáceas que no corresponden a la epidermis. Las glándulas sebáceas están compuestas de células llenas de lípidos que permiten la síntesis del sebo, una sustancia de hidratación de la piel para su protección. En la Ilustración 7, se ve una representación del infundíbulo y del istmo de un folículo piloso en una imagen histológica con tinción hematoxilina eosina. Entonces el infundíbulo se puede considerar parte de la epidermis aunque el resto del folículo piloso sea parte de la dermis.

Las glándulas sudoríparas son otras estructuras de la piel pero que no pertenecen a la epidermis. Son las glándulas que secretan el sudor. Estas estructuras son difíciles de segmentar para los patólogos ya que en caso de corte longitudinal como el de la Ilustración 7, distinguir la zona en la cual se termina el infundíbulo y empieza el istmo puede resultar en un reto.

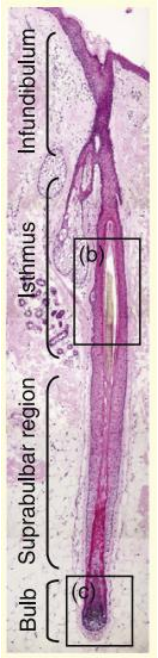


ILUSTRACIÓN 7: CORTE DE UN FOLÍCULO PILOSO EN IMAGEN HISTOLÓGICA H&E [14]

1.2.3. Las lesiones melanocíticas spitzoides

Las lesiones melanocíticas spitzoides pueden ser de varios tipos: melanoma spitzoide o nevus de Spitz. El primer tipo es maligno y corresponde a una forma de cáncer cuando el otro es benigno. Pero los dos tipos de lesiones son difíciles de diferenciar y diagnosticar para los dermatólogos y los patólogos [4]. Estos errores de diagnóstico pueden tener graves repercusiones en los pacientes que presenten un nevus de Spitz diagnosticado como melanoma maligno debido a que se les prescribirá erróneamente un tratamiento de quimioterapia o radioterapia que pondrá en peligro su salud y su calidad de vida. Por otra parte, una persona con melanoma malignado pero diagnosticado como nevus de Spitz puede perder un tiempo importante para dar la terapia adecuada a tiempo y maximizar su probabilidad de pronóstico favorable. El melanoma spitzoide puede ocurrir en niños, aunque es más frecuente en adultos. Corresponde a lesiones en la piel que cambian de tamaño. Una característica determinante para su diagnóstico es la proliferación melanocítica.

El melanoma spitzoides es un tipo de melanoma maligno pero muy similar a una lesión cutánea benigna. Se localiza con mayor frecuencia en la cabeza o en las extremidades. Es un tipo de melanoma bastante especifico que no sigue los criterios clásicos de caracterización de los melanomas de la Ilustración 8 y utilizados para el diagnóstico del melanoma.

ILUSTRACIÓN 8: REPRESENTACIÓN DEL SCORE ABCDE PARA EL MELANOMA [15]

Los melanoma invaden a veces la epidermis y resulta difícil aun para los patólogos diferenciar la epidermis de la dermis. En estos casos se usan otros marcadores de histoquímica para diferenciarlos. El melanoma spitzoide puede afectar a la epidermis y la dermis como una masa confluente por lo que la segmentación de la epidermis es útil en el apoyo digital a la patología. El melanoma spitzoide puede ocurrir de novo o dentro de una lesión nevus de Sptiz ya existente. Un nevo es cualquier lesión cutánea congénita de la piel.

Un nevus de Sptiz es un tipo benigno de lunar que suele ocurrir en la infancia [16], un ejemplo de este nevus se ve en la Ilustración 9. Se distingue de otros tipos de nevus por su tamaño más grande y el hecho de que suele crecer durante un periodo de tiempo y ese carácter incierto llega a la recomendación de su extirpación.

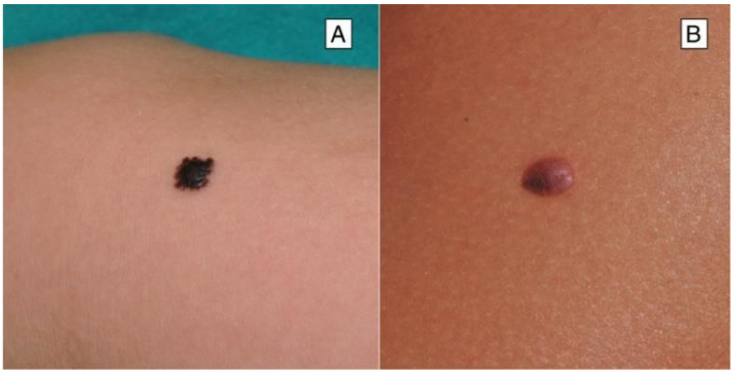


ILUSTRACIÓN 9: EJEMPLOS DE NEVO DE SPITZ [16]

El nevo puede ser de tres tipos como presentado en la Ilustración 10: en la unión epidérmica, intradérmico o compuesto que es una mezcla de los dos casos precedentes. Los nevos de unión consisten en nidos de células redondas que crecen a lo largo de la unión dermoepidérmica. Los nevos compuestos crecen en la dermis subyacente como nidos y los nevos intradérmicos crecen de forma parecida pero solo dentro de la dermis.

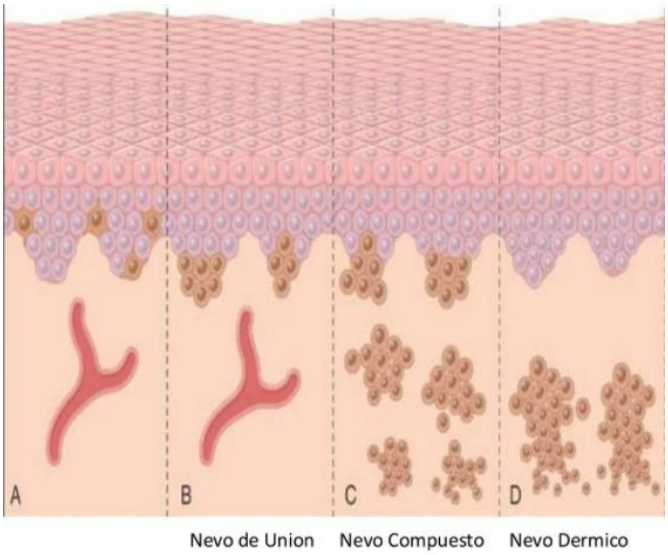


ILUSTRACIÓN 10: DIFERENTES TIPOS DE NEVOS [17]

El nevo melanocítico adquirido es la lesión más frecuente con la que se encuentran los patólogos y dermatopatólogos pediátricos en su práctica diaria. En la mayoría de los casos, hay pocas dificultades en el diagnóstico histopatológico. Sin embargo, es la lesión melanocítica adquirida conocida como nevus de Spitz, con sus características atípicas intrínsecas, la que se convierte en un reto, ya que hay muchos pasos parecidos del punto de vista histopatológico y biológico entre el tumor de Spitz atípico y el melanoma spitzoide. La frustración con algunas de estas lesiones spitzoides surge debido a que incluso los expertos en la materia no consiguen ponerse de acuerdo en cuanto a la diferenciación de una y otra, incluso a nivel de genética molecular.

1.2.4. El proyecto CLARIFY

Este trabajo se puede enmarcar en un proyecto de investigación europeo que coordina el CVBLab (Computer Vision & Behaviour Analysus Laboratory), laboratorio donde el estudiante ha realizado sus prácticas de empresa. Este proyecto se llama CLARIFY para inteligencia artificial en la nube para patología (*CLoud ARtificial Intelligence For pathologY* en inglés) [18]. Es un programa de investigación internacional de ingeniería y medicina. Tiene como objetivo desarrollar un entorno de diagnóstico digital automatizado basado en algoritmos de inteligencia artificial para facilitar la interpretación y el diagnóstico de WSIs. Se proponen algoritmos de datos orientados a despliegues en la nube para facilitar el acceso a la patología digital y maximizar sus beneficios. La plataforma en la nube permite un almacenamiento, una recuperación y una compartición mejor de la base de datos, mientras se haga de forma segura y asegurando la interoperabilidad de los datos. Este proyecto se dedica también a la formación de doce investigadores en inteligencia artificial. Se busca en este proyecto conseguir una mejora en el flujo de trabajo en los laboratorios de patología. El proyecto propone lograr un mejor intercambio de conocimientos y alcanzar una mejor toma de decisión en casos complejos de patología.

El proyecto CLARIFY se focaliza en tres tipos de cáncer específicos y desafiantes que sirven para poner a prueba la metodología desarrollada:

- Cáncer de mama triplemente negativo (TNBC)
- Cáncer de vejiga no invasivo de alto riesgo (HR-NMIBC)
- Lesiones melanocíticas spitzoides (SML).

Este trabajo es de interés para el estudio del último tipo de cáncer citado ya que se utilizan imágenes histopatológicas que presentan lesiones melanocíticas spitzoides como detallado en la sección anterior.

1.3. Estado del arte de la segmentación automática de la epidermis en imágenes histológicas

Los autores de los artículos de referencia de métodos de segmentación automática de la epidermis proponen técnicas en las cuales realizan previamente una segmentación gruesa de la epidermis. Esta primera segmentación se realiza según la intensidad de la tinción, luego se utilizan otros algoritmos para obtener resultados de segmentación de mayor precisión.

El método GTSA (*global thresholding and shape análisis*) presentado en el artículo [19] consiste en una segmentación empleando un umbral sobre el canal rojo de las muestras junto con el análisis de la forma. El canal monocromático extraído proporciona una información discriminante entre las áreas de la epidermis y de la dermis, lo que permite aplicar fácilmente el umbral. Se propone un enfoque multi resolución que en primer lugar realiza la segmentación de la imagen de baja resolución para luego analizar las imágenes de alta resolución a mano o de forma automática. Este método logra una sensibilidad de 0.92 y una especificidad de 0.97 en datos de validación. Se observa que la técnica propuesta ofrece un buen rendimiento usando la información discriminante del canal rojo de la imagen. Sin embargo, el articulo no provee la evaluación con la métrica de *dice* que se adapta mejor a la tarea especifica de la segmentación.

La técnica CET (*contrast enhancement and thresholding method*) mencionada en el artículo [20] es parecida al GTSA pero se añade una normalización del color de las imágenes para mejorar el contraste entre la epidermis y el resto de tejido epitelial. A continuación, se realiza un umbral global en la imagen mejorada con contraste, que se crea a partir de una combinación lineal de la imagen convertida en escala de grises de la WSI normalizada y del componente azul-amarillo de su imagen convertida en CIELAB. Esta técnica obtiene resultados experimentales altos con una especificidad media de 0.97 y una sensibilidad media de 0.96. El algoritmo es también capaz de segmentar la epidermis en imágenes que presentan diferentes niveles de daño tisular.

MCGT para *morphological closing and global thresholding-based technique* en inglés es la técnica desarrollada en el artículo [21] que presenta un sistema para medir la profundidad de la invasión del melanoma. En este artículo, se realiza la segmentación de la epidermis mediante la operación de cierre morfológico y una técnica basada en un umbral global sobre la imagen. Este método se basa en la hipótesis de que el cierre morfológico permite eliminar todos los pixeles detectados con el umbral global pero que están en la dermis como glándulas o núcleos de células. Pero la elección correcta del elemento estructurante para realizar estas operaciones morfológicas es una tarea compleja para que se adapte a la variedad de imágenes de biopsias de piel sin afectar a la calidad de la segmentación de la epidermis. El algoritmo propuesto segmenta la epidermis en imágenes en escala de grises y mide la intensidad de la melanina utilizando un clasificador SVM preentrenado. Para evaluar la calidad de los resultados del método propuesto, los autores comparan la máscara de epidermis computacionalmente obtenida con la anotada por tres patólogos. En 70% de los casos, el error entre el modelo y los médicos es de menos de 2%. Los resultados tienen un error promedio de 3,9 µm en la medición de la invasión del melanoma en el tejido en 40 imágenes de validación. La técnica permite una buena repetibilidad y reproducibilidad. El método tiene limitaciones cuando la zona de mayor intensidad es pequeña y en borde de la imagen. En este caso, el histograma de color es ineficaz y aleatorio.

Otra técnica presente en la literatura se basa en la medición del espesor de la epidermis y recibe el nombre de THM por sus siglas en inglés *thickness measurement* [22]. Los primeros pasos de este

método son similares a los de GTSA pero gracias a la segmentación con más alta resolución, operaciones de post procesado como una clasificación con el algoritmo *K-means* y basándose en la morfología con cierre, detección del esqueleto e identificación del eje principal es posible obtener buenos resultados de segmentación de la epidermis. El articulo propone una comparativa con los métodos propuestos en los artículos anteriores. El método THM supera en sus resultados las métricas obtenidas con las técnicas CET, GTSA y MCGT. Aunque las sensibilidades del método THM son ligeramente inferiores (90,39% en datos de entrenamiento y 92,78 % en datos de test) a las de la técnica GTSA (98,13% en datos de entrenamiento y 98,42 % en datos de test), la técnica THM logra presiones mayores de un 20% aproximadamente que la técnica GTSA. Las métricas de precisión de los métodos GTSA y CET son bajas (56,53% para el método CET y 75,01% para el método GTSA en datos de entrenamiento contra 98,69% para el método THM). Esto se debe al número alto de falsos positivos en las imágenes con muchos núcleos celulares en la dermis. La técnica MCGT segmenta asimismo muchas regiones de falso positivo en la dermis ya que el cierre morfológico no consigue eliminar las regiones de mayor intensidad en la dermis.

El articulo [23] propone una técnica basada en el análisis de la porosidad y el análisis de la concentración de la tinción (en adelante denominada técnica PASC para *porosity analysis and stain concentration analysis* en inglés) en WSI de múltiples fuentes muestreadas a 10x. Se obtiene una segmentación gruesa rellenando los espacios vacíos y realizando un umbral global sobre la densidad. El siguiente paso consiste en calcular la concentración de las tinciones de hematoxilina (H) y eosina (E) y rechazar la sangre, el estrato córneo y el colágeno denso realizando un umbral global en los mapas de concentración de E, H y H&E. Es un método robusto frente a las variaciones de la imagen, tanto en la tinción como en la iluminación, y no hace suposiciones sobre el tipo de trastorno de la piel. El método propuesto ofrece un rendimiento superior en comparación con las otras técnicas previamente citadas. La comparativa propuesta en el artículo muestra que el método PASC tiene una sensibilidad media de 0.87 contra 0.48 para THM, 0.86 para GTSA y 0.99 para CET sobre los datos de validación. La especificidad obtenida por los modelos GTSA, THM y PASC es similar con los valores 0.90; 0.99 y 0,95 respectivamente. La técnica PASC mejora la precisión con un valor de 0.57 comparativamente con la técnica de THM (0.55), la técnica GTSA (0.44) y la técnica CET (0.21). El método PASC presenta limitaciones para imágenes con una infiltración abundante de núcleos en la dermis como los métodos anteriores. Los apéndices de la piel como los folículos pilosos o las glándulas sebáceas limitan también la precisión de la segmentación obtenida. El alto rendimiento de esta técnica permite que sea usada como paso previo a la detección de melanocitos o al diagnóstico de melanoma por ejemplo.

No obstante, la mayoría de estos métodos realizan suposiciones sobre la uniformidad de la tinción de las imágenes del conjunto de datos y el contraste suficiente entre la epidermis y la dermis. Estas hipótesis no permiten modelar la complejidad de la variabilidad de las imágenes de piel posible en la práctica clínica sobre todo en casos patológicos. Estos métodos se basan en esencia sobre el uso de un umbral global de la imagen considerando que las estructuras de más intensidad dentro de la dermis se pueden quitar luego con post procesado. Pero estos métodos puede que no tengan la intensidad o la forma adecuada para que las etapas de post procesado las eliminen o que no sea adecuado eliminarlas si son infundíbulo de folículos pilosos que pertenecen a la epidermis. Además, estos métodos requieren la extracción o selección de características dentro de las imágenes que es una tarea compleja y puede introducir sesgo en el modelo. Los resultados obtenidos con los métodos descritos anteriormente presentan muchas regiones de falsos positivos dentro de la dermis.

En 2019 se publicó el artículo [24] en el que se lleva a cabo la tarea de segmentación mediante una red neuronal convolucional de arquitectura U-Net obteniendo buenos resultados. Se entrena la red con *patches* de 512*512 extraídos de las WSI teniendo en cuenta el desbalance de clase. El sobre muestreo de la epidermis con submuestreo del fondo y de los otros tejidos permite paliar a este problema, pero también aumenta el riesgo de *overfitting* ya que se repite mucho la misma información

de epidermis. Este articulo usa el *data augmentation* con rotación y dando la vuelta a la imagen para mejorar los resultados obtenidos.

Usan también ciertas etapas de post procesado de alisado con un *kernel* de promediado y un umbral sobre la imagen con el método de Otsu [25]. Estas etapas permiten obtener una segmentación buena de la epidermis en imágenes histopatológicas. Este método es el que hoy en día presenta mejores métricas. Es el más novedoso y que mejora (o iguala en ciertos casos) los resultados de segmentación de las técnicas previamente comentadas.

Para los datos de entrenamiento de los diferentes métodos presentados, se obtiene una sensibilidad media de 0.39 para el método THM, 0.55 para el método GTSA, 0. 89 para el método PASC, 0.97 para el método CET y 0.97 para el método U-Net. En esta métrica, los métodos CET y U-Net son los que dan mejores resultados pero en las métricas de *Positive Predictive Value*, *Dice* score y *Matthews Correlation Coefficient*, la U-Net obtiene métricas de 0.85 en promedio cuando los otros métodos no superan 0.51 sobre los mismo datos. El método del artículo [24] se presenta entonces como la técnica que permite los mejores resultados en cuanto a sensibilidad, especificidad, PPV, *dice* y *Matthews Correlation Coefficient* de la literatura. Mejora significativamente las técnicas previas en cuanto al dice. El dice de esta técnica es de 0.89 en datos de test cuando vale 0.47 para el método CET; 0.39 para el método GTSA; 0.45 para el método THM y 0.68 para el método PASC sobre los mismos datos. Al nivel cualitativo, la segmentación obtenida iguala las anotaciones realizadas por los patólogos en la mayoría de los casos. En cambio, ciertas imágenes siguen siendo difíciles de analizar con esta técnica y la máscara de anotación obtenida presenta falsos negativos o positivos.

Además, este último método presenta etapas de sobre muestreo/submuestreo que tienen un riesgo de introducir cierto sesgo en el método obtenido. Las etapas de post procesado también son propias de una base de datos y no son idóneas para adaptarse a la variabilidad de los datos de la práctica clínica diaria. El articulo destaca como limite el tiempo de computación para la inferencia de una nueva imagen.

Este trabajo final de master propone un nuevo método de segmentación automática de la epidermis en imágenes histopatológicas de piel con melanoma spitzoide. La técnica propuesta se basa también en una arquitectura de U-Net pero añadiendo bloques residuales para mejorar los resultados obtenidos. Este enfoque propone usar la métrica de *dice* ponderado para manejar el desequilibrio de clases y dos redes encoder-decoder a dos niveles de resolución para obtener una segmentación de precisión de la epidermis sin necesidad de etapas de post procesado. Propone también realizar las etapas con imágenes de alta resolución solo con zonas detectadas como epidermis potencial y así quitar las regiones no indicadas y mejorar el tiempo de computación de la inferencia del algoritmo. Según se conoce, este trabajo supone el primer trabajo que propone un método específico para biopsias de melanoma spitzoide.

La U-Net del articulo inicial toma como imagen de entrada una imagen de tamaño especificado y entrena a esta resolución elegida. En el caso de la segmentación de la epidermis, la red se entrena sobre la imagen completa (es la WSI pero con una resolución menor). Pero la epidermis como etiqueta de objetivo esta menos representada que el fondo. Así que puede ser de interés de entrenar a una resolución más baja. Hay artículos que se interesan a redes de multi-resolución para segmentación semántica en WSI. El articulo [26] presenta dos métodos de multi-resolución basados sobre la arquitectura de la U-Net que tienen resultados mejores que la U-Net a una sola resolución.

La información del vecindario de las estructuras de interés es muy importante para que el patólogo pueda clasificar la región. Pero para considerar esta información alrededor de las estructuras de interés se necesita hacer *patches* de la WSI. Los *patches* son segmentos pequeños de la WSI. La utilización de

patches es muy común en *Deep learning*. El modelo entrenado predice cada *patch* individualmente y luego se unen las predicciones obtenidas para formar la predicción de la imagen completa. Esos *patches* tienen que ser los más grandes posibles para considerar más información. Mientras que si se preserva muy bien la calidad de la imagen eso implica una imagen grande en entrada de la red que alcanza los limites computacionales del servidor.

1.4. Objetivos del trabajo

El objetivo de este trabajo es integrar los conocimientos en inteligencia artificial y tratamiento de imágenes médicas del estudiante para llevar a cabo un modelo automático de segmentación de la epidermis en imágenes histopatológicas de biopsias de piel con lesiones melanocíticas. Este trabajo se integra en el proyecto hacia la patología digital de un laboratorio de investigación focalizado en la inteligencia artificial para resolver retos de análisis de datos.

Se implementarán varias arquitecturas de redes neuronales convolucionales a diferentes resoluciones para detectar de forma más precisa y robuste posible la capa superior de la piel en las imágenes con tinción hematoxilina eosina.

Los objetivos específicos que se deben abordar para lograr el objetivo principal presentado son los siguientes:

a. Estado del arte de la literatura científica sobre la segmentación de la epidermis y de forma más general sobre la segmentación en imágenes médicas.
b. Estado del arte de las arquitecturas de redes neuronales para la segmentación.
c. Creación de la base de imágenes disponibles con sus máscaras de anotación correspondiente (CLARIFYv1) y segmentación manual de la segunda base (CLARIFYv2) bajo la supervisión de un patólogo.
d. Análisis de los requerimientos computacionales y selección del entorno de programación.
e. Implementación de dos arquitecturas de redes convolucionales a dos resoluciones.
f. Validación de los modelos predictivos generados a nivel de *patch* y comparación de los resultados de las arquitecturas de red neuronal convolucional profunda utilizadas sobre la base de datos CLARIFYv1.
g. Reconstrucción de las imágenes y evaluación del modelo a nivel de imagen completa sobre la base de datos CLARIFYv1.
h. Validación del modelo utilizando la base de datos CLARIFYv2 a nivel de imágenes completas y de *patches.*

2. Marco teórico

En esta parte se van a presentar los conceptos teóricos necesarios para la comprensión de este trabajo.

2.1. Introducción

El *Machine Learning* es una parte de la inteligencia artificial que se refiere a la capacidad de un modelo informático a aprender a realizar una tarea sobre datos sin haber sido programado de forma explícita para resolverla. Se basa en encontrar patrones en los datos para realizar predicciones sobre datos nuevos. Entonces intrínsecamente un algoritmo de *Machine Learning* funciona con datos. Se extraen características relevantes de los datos según el tipo de estudio como la media, desviación estándar (*standard deviation*, abreviado en std en este trabajo) y energía del histograma de la imagen. Estos datos se ponen en entradas del algoritmo elegido. Existen dos grandes tipos de algoritmos de *Machine Learning*: los algoritmos supervisados y no supervisados.

Un algoritmo supervisado tiene en entrada datos y una etiqueta (*label* en inglés) que corresponde a la salida deseada que el modelo debe encontrar realizando sus predicciones. En este caso el modelo necesita un entrenamiento, lo que significa una etapa en la cual, según sus datos de entrada, la salida que obtiene y la comparación con la etiqueta, los parámetros del algoritmo se ajustan para obtener una salida lo más parecidas posible a la etiqueta. Una función permite comparar la salida de predicción y la etiqueta, esta recibe el nombre de función de pérdidas o de coste. La función de pérdidas refleja los errores de predicción del modelo comparativamente con la etiqueta y por tanto, el objetivo es minimizarla lo máximo posible. Se pueden añadir métricas que miden positivamente hasta qué punto la predicción y la etiqueta se parecen. Se trata de maximizar las métricas en cuestión. Las métricas suelen ser bastante específicas a la tarea de *Deep learning*.

Un algoritmo no supervisado solo intenta sacar los mejores resultados posibles a partir de los datos no etiquetados. Entonces al revés del tipo de modelo últimamente presentado no puede tener una función que le repercuta según los errores que hace ya que los datos no son etiquetados. Este tipo de algoritmos se usa por ejemplo en tareas de clustering u optimización donde la mejor forma de hacerlo no es previamente conocida.

En el caso de una clasificación binaria, los datos pueden tener o bien el valor 1 que se puede llamar valor positivo o bien el valor 0 que se puede llamar valor negativo según el objetivo del estudio. Por ejemplo, se puede realizar un modelo que predice si una muestra es de un tejido sano o patológico, en este caso el valor positivo corresponde al tejido patológico porque es sobre todo el que se quiere detectar. El tejido sano en cambio se verá atribuido el valor negativo.

El número de predicciones positivas que se corresponden a valores positivos de la etiqueta o *ground truth* se llama verdaderos positivos o true positive en inglés y se suele notar TP. De la misma manera el número de predicciones negativas que se corresponden a valores negativos del *ground truth* se llama verdaderos negativos notados TN para true negative en inglés. Unas tasas de TP y TN altas significan que la predicción se parece a la etiqueta de *ground truth*, lo que es el objetivo deseado. Al contrario, las predicciones positivas pero que no lo son en la etiqueta de *ground truth* son llamados falsos positivos (FP en inglés) y las predicciones negativas pero que son positivos en la etiqueta de *ground truth* son los falsos negativos (FN en inglés). Estas abreviaciones se van a usar a lo largo de este trabajo en las fórmulas correspondientes.

En este caso la matriz de confusión es la matriz compuesta de estos valores de la forma siguiente:

TABLA 1: MATRIZ DE CONFUSIÓN

		Etiqueta de referencia (*ground truth*, en inglés)	
		1	0
Predicción del clasificador	1	TP	FP
	0	FN	TN

Los modelos de predicción y otras tareas de análisis de datos (optimización, clasificación, selección de características…) pueden tener diferentes arquitecturas de algoritmos como el perceptrón detallado en la sección siguiente.

Se dice que una red neuronal es profunda en función del número de capas que la componen. Cuando este número es muy alto el modelo se puede enmarcar dentro del paradigma del *Deep Learning*. Significa que aprende con profundidad los datos sin que se pueda saber exactamente cuáles son todos los pasos de una neurona a otra para obtener la salida. Son modelos de caja negra.

El *Deep learning* comparativamente con el *Machine learning* tiene la ventaja de no apoyarse en características extraídas a mano de las imágenes. Los algoritmos de aprendizaje profundo pueden aprender el modelo directamente desde el *raw data* que son los datos brutos no modificados. También presentan la ventaja de ser bastante fáciles de poner en aplicación y de ser de alta precisión. Esta alta precisión se suele obtener a pesar de una mayor coste computacional del entrenamiento del modelo. El *Deep learning* se puede aplicar a la visión artificial en el campo de la ingeniería biomédica.

La visión artificial es un campo científico de visualización de imágenes por ordenadores que consiste en la adquisición, procesamiento y análisis de imágenes del mundo real. Proporciona información tratada por ordenador y tiene muchos campos de aplicación. Se busca que los ordenadores realicen un tratamiento de la información visual parecida a la que es capaz de realizar el ojo y cerebro humano. Entonces el ordenador debe ser capaz de detectar patrones de interés en las imágenes y de proporcionar una salida visual de los resultados obtenidos. En el caso de imágenes médicas, la visión artificial puede permitir el diagnóstico, pronóstico o predicción de la evolución de patologías con las múltiples fuentes de imagen como RMI, rayos X, OCT entre otras.

Las tareas que se pueden llevar a cabo con *Deep learning* en visión artificial pueden ser de clasificación de imágenes por ejemplo, pero en este trabajo es la tarea de segmentación que se va a presentar con más detalle.

La tarea de segmentación consiste a determinar dentro de una imagen zonas de interés creando una máscara que indica la posición dentro de la imagen de la estructura. La segmentación de imágenes es un problema fundamental en visión artificial ya que consiste en dividir las imágenes en múltiples segmentos y objetos. Presenta una gama amplia de aplicaciones como en vehículos autónomos para la detección de peatones o en el análisis de imágenes médicas con la extracción de tumores y medición de volúmenes de tejidos. Puede también servir de etapa previa a un análisis posterior reduciendo la información útil a objetos detectados, por ejemplo después de la segmentación de la zona tumoral un médico o un algoritmo de clasificación pueden calificar el tumor de benigno, maligno y ayudar en la predicción de su futura evolución en el organismo del paciente. La Ilustración 11 muestra un ejemplo de varios tipos de tejidos segmentados dentro del cerebro. La segmentación realizada permite dar sentido a los objetos detectados en las imágenes analizadas y

delimitar las zonas de lesiones en imágenes de resonancia magnética que es crítico a la hora de realizar radioterapia o una cirugía de terapia.

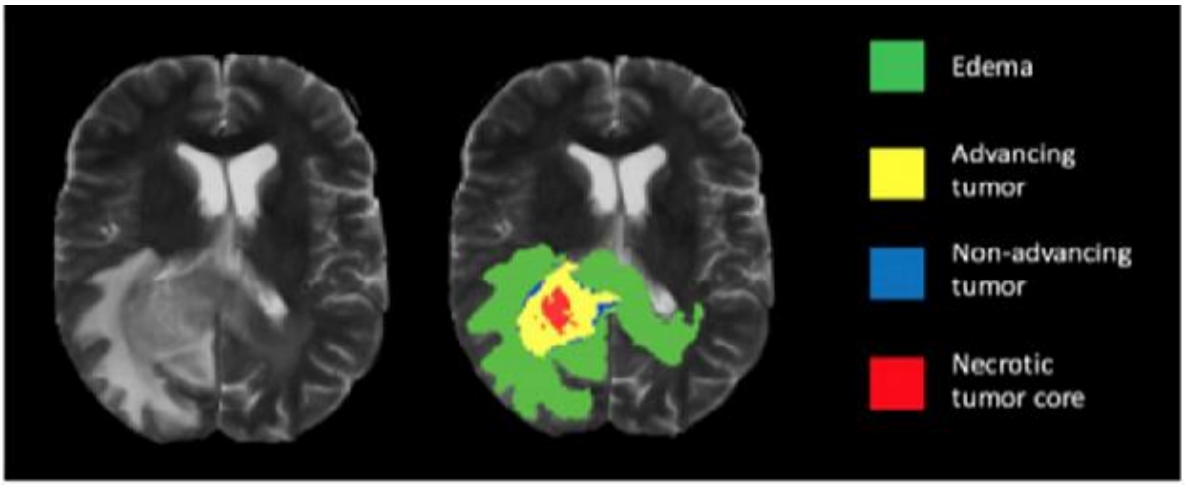


ILUSTRACIÓN 11: EJEMPLO DE SEGMENTACIÓN EN IMÁGENES IRM DEL CEREBRO [27]

La segmentación de imágenes médicas se puede aplicar a muchos tipos de imágenes como las imágenes histopatológicas que son las estudiadas en este trabajo. La Ilustración 12 propone un ejemplo de segmentación semántica de núcleos en imágenes de histología de pulmón mediante una red neuronal profunda. El articulo fuente propone un guion de la introducción de la tecnología de segmentación automática en el flujo de trabajo del laboratorio de patología. Se ve en la ilustración la máscara aquí con cuatro etiquetas que se desea obtener cuando se programa un algoritmo de segmentación automática. La máscara es del tamaño de la imagen y representa con colores las zonas delimitadas de los pixeles que corresponden a un tipo celular en concreto. La máscara de segmentación da información del significado dentro de la imagen mientras teniendo un peso computacional menor que la imagen completa con todos los niveles de colores correspondientes.

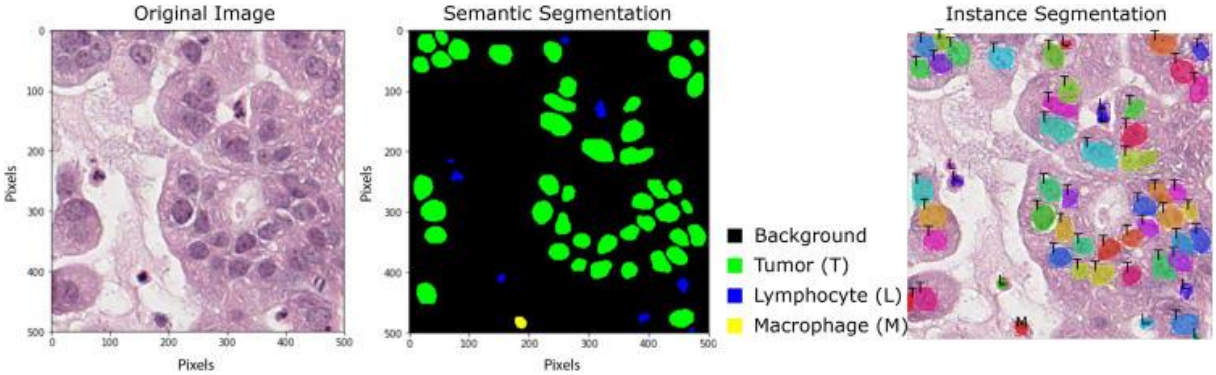


ILUSTRACIÓN 12: EJEMPLO DE SEGMENTACIÓN DE NÚCLEOS DE DIFERENTES CÉLULAS EN IMÁGENES H&E DE PULMÓN [28]

2.2. Redes neuronales artificiales

El concepto de red neuronal artificial utiliza la idea de las redes neuronales biológicas compuestas de unidades llamadas neuronas que se conectan a otras formando conexiones especializadas. En la biología del sistema nervioso, una neurona tiene dendritas que sirven como puertas de entradas activadas por la liberación de neurotransmisores de los axones de neuronas vecinas como se ve en la Ilustración 13. En el soma de la neurona se procesa la información que llega como impulso eléctrico y según su intensidad y su frecuencia la neurona se activa y propaga un impulso a través de su axón hasta el botón sináptico que se conecta con la neurona siguiente.

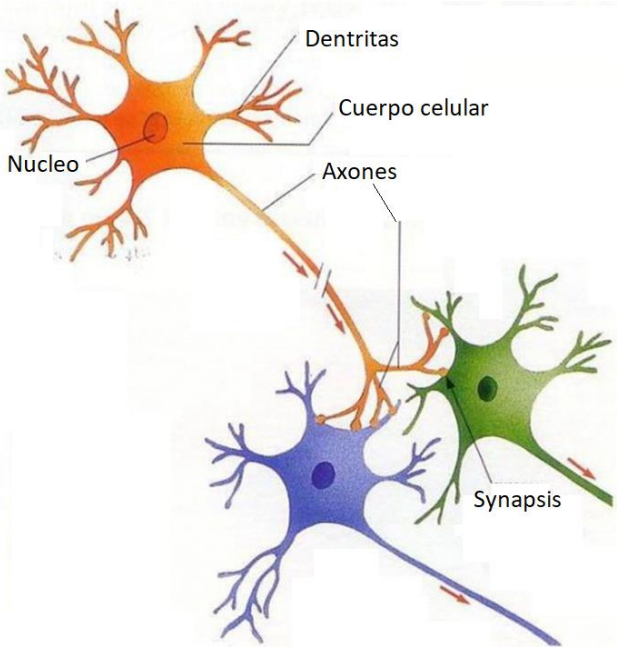


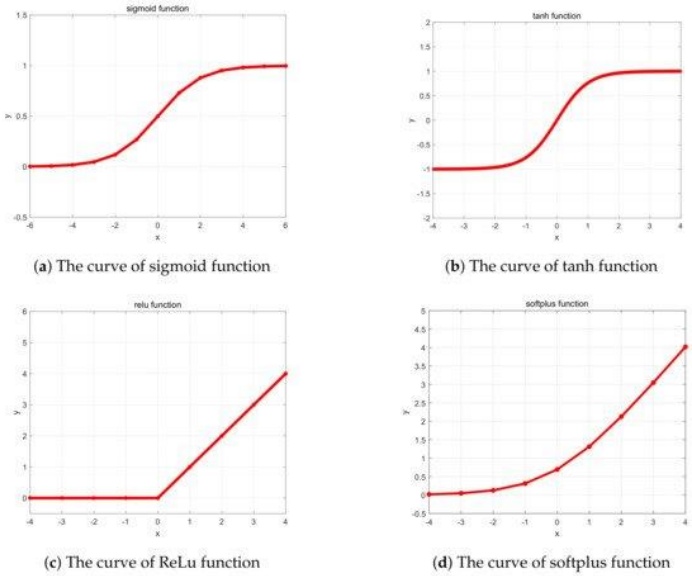
ILUSTRACIÓN 13: CONEXIÓN ENTRE NEURONAS [29]

Este proceso biológico ha inspirado un modelo computacional llamado red neuronal artificial. La programación de este tipo de algoritmo pretende imitar el aprendizaje que se ha realizado en el sistema nervioso a nivel de la cantidad de neurotransmisores que activan las dendritas de las neuronas. Creando un aprendizaje similar pero a nivel informático basándose en datos permite ser capaz de extraer características y patrones de los datos de entrada para realizar tareas complejas de clasificación, segmentación, predicción u optimización. Al igual que el sistema nervioso central humano compuesto de muchas neuronas, una red neuronal artificial debe tener muchas neuronas y muchas capas, aunque eso representa un reto computacional. Hoy en día, la tecnología ha avanzado mucho para permitir la realización de algoritmos tan grandes y complejos y ha permitido llevar a cabo tareas que antes se descartaban por la poca viabilidad que ofrecía su coste computacional.

Una red neuronal artificial está compuesta de unidades básicas que son las neuronas, ya comentadas, que se conectan a otras. Los enlaces que conectan una neurona a otra pueden tener pesos que incrementan o inhiben el estado de activación de la neurona siguiente según el valor de la salida de la primera neurona. La salida de la neurona debe sobrepasar un umbral para que la información se propaga a la neurona siguiente. La función de umbral que se utiliza se llama la función de activación.

Existen varios tipos de función de activación, se suelen dividir entre las funciones de activación lineales y las no lineales. La ventaja de usar una función de activación no lineal es que permite introducir la no linealidad en un perceptrón clásico que tiene sus otros parámetros con relaciones

únicamente lineales. Como función de activación muy utilizada se pueden citar la función sigmoide, la tangente hiperbólica, la función ReLu y la función *softplus* como ploteadas en la Ilustración 14.



(a) The curve of sigmoid function (b) The curve of tanh function

(c) The curve of ReLu function (d) The curve of softplus function

ILUSTRACIÓN 14: GRÁFICOS DE 4 TIPOS IMPORTANTES DE FUNCIÓN DE ACTIVACIÓN [30]

La función sigmoide fue empleada en las primeras redes neuronales por su consistencia con las sinapsis de las neuronas del sistema nervioso y su derivada que es fácil de obtener. Se usa poco hoy en día debido a su tendencia a disminuir mucho el gradiente cuando la pendiente de la función se acerca de cero, lo que hace difícil el entrenamiento de los pesos de la red y afecta a la tasa de convergencia del modelo. Su fórmula es: $f(x) = \frac{1}{1+e^{-x}}$

La función de tangente hiperbólica es una actualización de la función sigmoide, tiene como la precedente una salida acotada pero su tasa de convergencia es mayor. Todavía esta función presenta el problema de la difusión del gradiente. Su fórmula es: $f(x) = \frac{1-e^{-2x}}{1+e^{-2x}}$

La función que actualmente más se una es la función ReLu que se compone de dos trozos lineales. Esta función hace que la salida sea cero para todas las entradas negativas e igual a la entrada para las entradas positivas. Forzar que algunos datos valgan cero permite activar menos neuronas y entonces proporciona una velocidad de cálculo más rápida. Estas ventajas hacen que se prefiere usar esta función comparativamente con las dos precedentes. Pero la función ReLu presenta también unos límites como un valor demasiado alto cuando un gradiente grande pasa por la función. Su fórmula es: $f(x) = max(x, 0)$. Existen también variaciones de esta función para adaptarse mejor al algoritmo de red neuronal según los casos y paliar a sus limitaciones.

La función *softplus* es parecida a la función ReLu pero presenta una curva ausente en la función ReLu. No pone a cero todas las entradas negativas sino a valores cercanos de cero. Su fórmula es: $f(x) = ln(1 + e^x)$. Esta función se usa menos por su coste computacional para calcularla, la función ReLu presenta la otra ventaja que se calcula muy fácilmente.

Los valores de los pesos se actualizan de manera a reducir la función de pérdida presentada en la sección 2.1 sobre el *machine learning*. Uno de los primeros algoritmo de red neuronal artificial fue el perceptrón inventado en 1957 y que presenta la base de una neurona con sus entradas, sus pesos y su función de activación para calcular una salida como se ilustra en la Ilustración 15. A continuación se presenta más en detalle el funcionamiento del algoritmo del perceptrón para entender su

generalización a un algoritmo de muchas neuronas organizadas en capas y basadas en la unidad de un perceptrón.

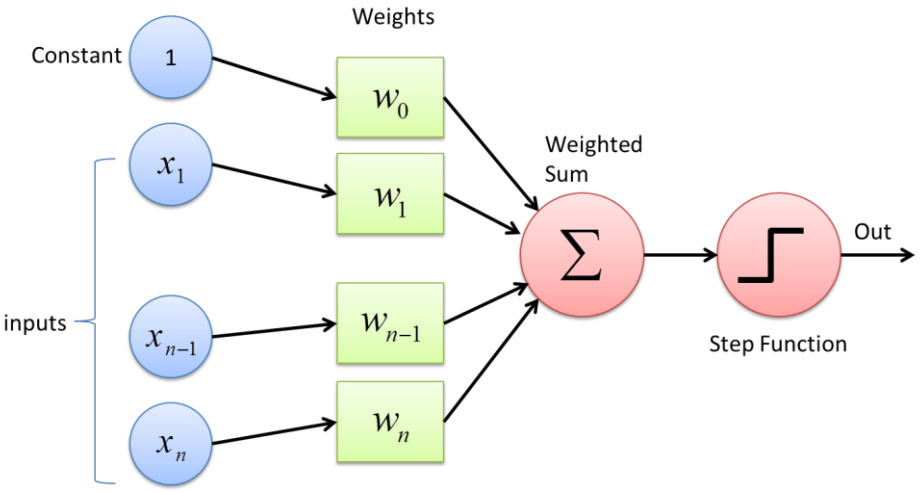


ILUSTRACIÓN 15: ESQUEMA DE LA ARQUITECTURA DE UN PERCEPTRÓN [31]

El perceptrón simple es un clasificador lineal supervisado. Considerando de las clases son linealmente separables con una dimensión previamente reducida, se busca el hiperplano de separación optimo entre dos clases. Si se conoce un conjunto de instancias x^k perteneciendo a la clase k y otro $x^{\bar{k}}$ no perteneciendo, este conjunto se llama conjunto de entrenamiento y se escribe como:

$$(x^1, y^1), (x^2, y^2), \dots, (x^m, y^m)$$

$$donde\ y^k = f(x^k)$$

$$\boldsymbol{f(x) = sgn(w.x) = \begin{cases} 1\ si\ w.x > 0 \\ -1\ si\ w.x \leq 0 \end{cases}}$$

ECUACIÓN 1: PERCEPTRÓN SIMPLE

El algoritmo del perceptrón consiste en inicializar aleatoriamente w con componentes aleatorias entre 0 y 1, luego para cada iteración t y cada k entre 1 y m, se calcula $z^k(t) = sgn(w^k(t).x^k)$ y si $z^k(t) \neq y^k$ se modifican los pesos w como

$w_i(t+1) = w_i(t) + \eta(z^k(t) - y^k)x_i^k$ con η el *learning rate* mientras ninguna componente del vector de pesos w haya cambiado en las p últimas iteraciones. Si las clases son linealmente separables este algoritmo converge a una solución en un número finito de iteraciones.

El perceptrón simple sólo clasifica patrones linealmente separables, existen sin embargo problemas que no son solubles ni siquiera usando como características funciones complejas de las originales. Para resolver aquellos casos en los que las fronteras de decisión no son hiperplanos sino funciones arbitrarias pueden usarse diversos modelos de redes neuronales como el perceptrón multicapa. Se considera el perceptrón como una célula básica de computación, llamada neurona artificial que toma n entradas (cada una de las componentes de una instancia x) y genera una salida 1 si pertenece a la clase aprendida y -1 sino.

$$s(x) = f\left(\sum_{i=0}^{n} w_i x_i\right)$$

donde f es la función signo, pero también puede sustituirse por cualquier función de tipo sigmoidal. Un perceptrón multicapa se construye con al menos dos capas de perceptrones, la primera

oculta con P superior a uno unidades y la segunda de salida con al menos una unidad donde las salidas de cada capa oculta actúan como entradas a la siguiente (detallado en la sección 2.3). La capa de entrada se limita a copiar el valor de x_i en la entrada correspondiente de las neuronas de la primera capa oculta. El problema está en encontrar los valores de los pesos w_i de cada neurona que hacen que la salida sea la requerida para cada patrón. Para esto existe el algoritmo de retro propagación (*backpropagation*). No obstante, este y todos los algoritmos propuestos tienen el problema de la caída en mínimos locales, es decir, pueden converger a un conjunto de valores de los pesos que no clasifica bien, dependiendo de los valores con los cuales se inicializaron. Si se usan varias neuronas en la capa de salida se puede implementar un clasificador multiclase.

2.3. Red neuronal convolucional

Una red neuronal convolucional es un tipo de perceptrón multicapa con una regulación que aprovecha patrones en los datos de complejidad creciente utilizando filtros a diferentes niveles de resolución. Las redes convolucionales se inspiran del proceso biológico entre neuronas. Durante la fase de entrenamiento la red optimiza sus filtros ajustan los pesos según las imágenes de entradas con su etiqueta. Así este tipo de red se libera del diseño con conocimientos previos e intervención humana de la extracción de características.

Convolucional neural networks (CNN) son redes neuronales basadas en la convolución. Una red neuronal convolucional está compuesta de diferentes capas que pueden ser:

- Capas convolucionales,
- Capas de *pooling*,
- Capas de activación,
- Capas completamente conectadas (conocidas como *fully connected*, FC en inglés).

La convolución en imágenes es una operación matemática que consiste en aplicar la multiplicación de cada pixel de la imagen por una matriz de menor tamaño llamada *kernel* y obtener una nueva imagen de resultado como explicado en [32]. El *kernel* suele ser de tamaño 3*3 o 5*5 pero otros tamaños son también posibles. La convolución coloca las imágenes de entrada a través de un conjunto de filtros convolucionales (*kernel*), cada uno de los cuales activa ciertas características de las imágenes. La ilustración 6 muestra un ejemplo de convolución discreta de una matriz 5*5 con un *kernel* 3*3. Las capas de convolución suelen ser 2D con una conectividad local que puede ser definida.

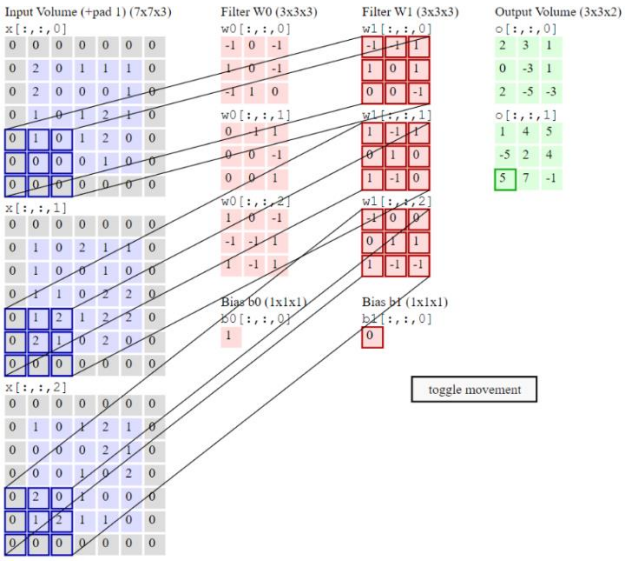


ILUSTRACIÓN 16: EJEMPLO DE CONVOLUCIÓN DISCRETA [33]

Las capas de *pooling* sirven para reducir el tamaño de la imagen. Simplifica la salida al realizar un submuestreo, lo que reduce la cantidad de parámetros que la red necesita conocer. El *max-pooling* consiste en conservar solo el pixel máximo entre varios vecinos como se ve en la ilustración 7.

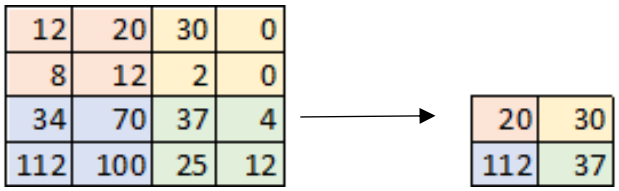


ILUSTRACIÓN 17: EJEMPLO DE *MAX-POOLING* (ILUSTRACIÓN PROPIA)

La capa de activación introduce la no-linealidad necesaria al modelo. El *rectified linear unit* (ReLu) permite un entrenamiento más rápido y efectivo al asignar valores negativos a cero y mantener valores positivos como se ve en la Ilustración 18. Es la función utilizada en vez de la función signo presentada en la arquitectura del perceptrón.

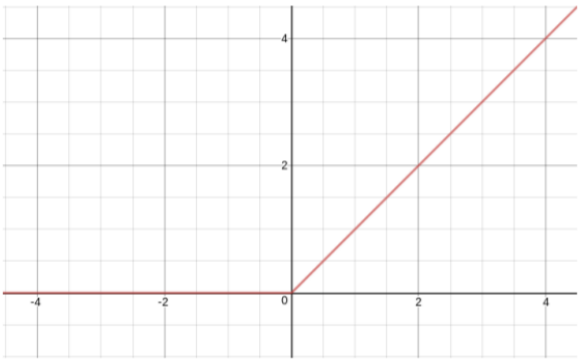


ILUSTRACIÓN 18: FUNCIÓN DE ACTIVACIÓN *RECTIFIED LINEAR UNIT* (RELU) [34]

Las operaciones de convolución, *pooling* y ReLu se repiten en las distintas capas y cada una aprende a detectar características diferentes. En estas capas se realiza la detección de características. Después de eso, la arquitectura de una CNN cambia a clasificación. La capa después de la última capa totalmente conectada (capa en rojo en la ilustración 9) genera un vector de K dimensiones donde K es el número de clases que la red podrá predecir. Este vector contiene las probabilidades para cada clase de cualquier imagen que se clasifica. La capa final de la arquitectura CNN utiliza una función *softmax* para proporcionar la salida de clasificación.

La función *softmax* se define como

$$\boldsymbol{p_k(x) = exp(a_k(x)) / \left(\sum_{k'=1}^{K} exp(a_{k'}(x)) \right)}$$

ECUACIÓN 2: SOFTMAX

Con $a_k(x)$ es la activación del canal k en el pixel de posición x. K es el número de clases y $p_k(x)$ es la aproximación de la función máximum. Entonces $p_k(x)$ tiene un valor cerca de 1 para el k que tiene $a_k(x)$ máximum y $p_k(x)$ tiene un valor cerca de 0 para los otros k.

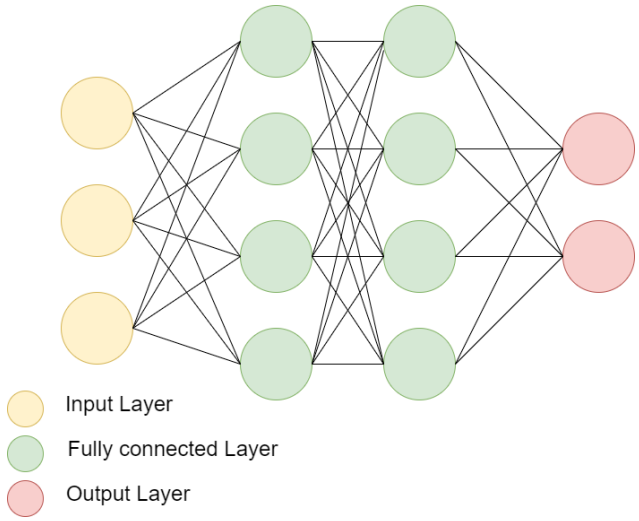


ILUSTRACIÓN 19: ESQUEMA DE CAPAS COMPLETAMENTE CONECTADAS (ILUSTRACIÓN PROPIA)

La Ilustración 20 presenta un ejemplo de la arquitectura global de una red neuronal convolucional con las capas mencionadas de convolución, *max-pooling* y *fully-connected*.

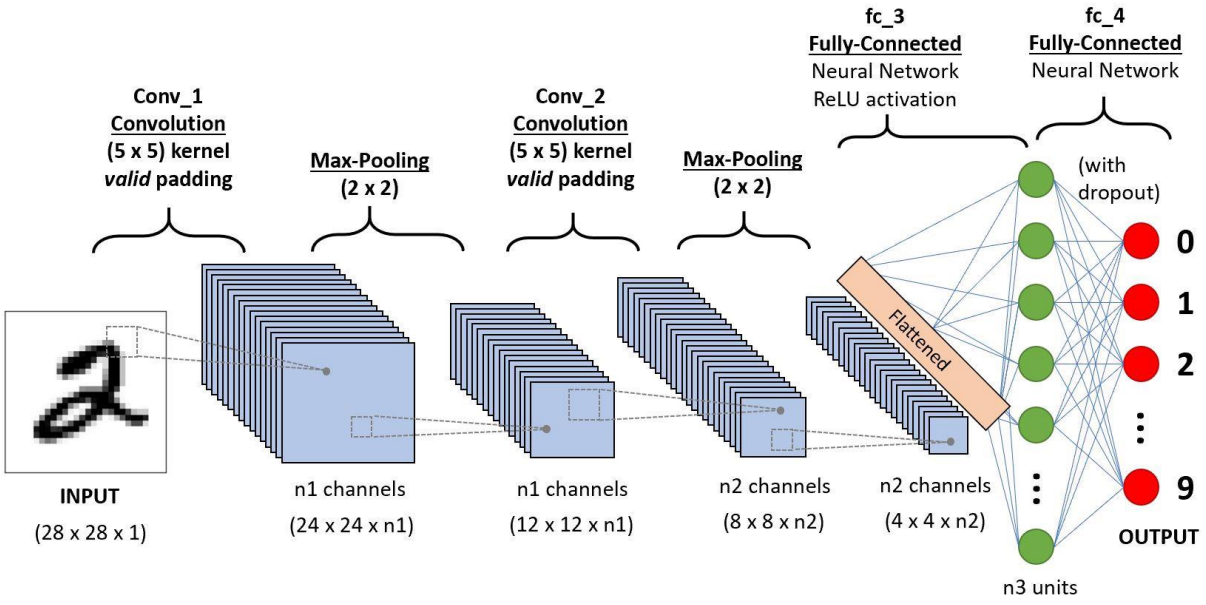


ILUSTRACIÓN 20: ESQUEMA DE UNA CNN [35]

La arquitectura de CNN tiene un uso amplio en imágenes médicas para la investigación y su aplicación en varios campos de la medicina. Puede servir para la predicción del diagnóstico de la diabetes mellitus como presentado en el artículo [36]. Sirve también en radiología que es un campo por esencia que trata con muchas imágenes con el objetivo de aumentar el rendimiento de los radiólogos y optimizar la atención al paciente. Puede presentar unos retos con conjunto de datos pequeños y sobreajuste como explicado en el artículo [37].

En el diagnóstico médico, la tarea de segmentación de imágenes es muy importante y también una tarea esencial en visión artificial. Comparativamente con tareas de clasificación, por ejemplo, la tarea de segmentación necesita información de bajo nivel lo que la hace más compleja. Las redes neuronales convolucionales profundas son muy utilizadas para llevar a cabo la segmentación semántica y de instancias en imágenes [38]. Dentro de las CNN, destacan muchos algoritmos que son

variaciones de redes convolucionales y se adaptan a tareas más específica. A continuación, se presenta uno de ellos que es la U-Net.

Las redes neuronales convolucionales ya se han visto usadas en el contexto del diagnosis de melanoma como en los artículos siguientes [39] y [40]. En el primer artículo, usan una arquitectura Mask-R-CNN como algoritmo de segmentación de regiones de potencial lesión melanocítica. El modelo obtenido puede ser entrenado con etiquetas ruidosas así que es bastante robusto y también permite limitar el tiempo de la anotación manual por un especialista de cada imagen. El segundo articulo realiza una segmentación semi-supervisada de células de nevus con una arquitectura de *autoencoder* con dos pasos de aprendizaje. Ambas técnicas presentan buenos resultados.

Para realizar el entrenamiento de una red neuronal con una base de datos importante se debe fijar el valor del *batch size*. El *batch size* es el número de muestras del conjunto de entrenamiento utilizada a cada iteración. Este número puede ser igual al tamaño del dataset completo (*batch mode*), en este caso la iteración corresponde a una época, puede ser pequeño pero no igual a uno (*mini-batch mode*) y puede ser igual a uno (*stochastic mode*). En el caso que el *batch size* vale uno, el gradiente y los parámetros de la red neuronal se actualizan después de cada muestra. Entonces, el *batch size* es un hiperparámetro importante que influye en la dinámica del algoritmo de aprendizaje. Este parámetro influye sobre la calidad de la estimación del gradiente de error entre la predicción del modelo y la etiqueta de *ground truth*. Como es una estimación estadística, cuantos más ejemplos de entrenamiento, más precisa y probable es la actualización de los pesos de la red. En contrario, un número reducido de muestra da lugar a una estimación que depende en gran medida de los ejemplos de entrenamiento específicos utilizados y puede resultar ruidosa. No obstante, estas actualizaciones ruidosas pueden dar lugar a un aprendizaje más rápido y, en ocasiones, a un modelo más robusto. Por lo tanto, es un parámetro que se debe ajustar cuidosamente según el problema y los datos en cuestión. Se suele usar *batch size* pequeños porque ofrecen un efecto de regularización al modelo, facilitan que un lote de datos de entrenamiento quepa en la memoria y porque suelen funcionar bien en casos genéricos.

Otros parámetros de interés a especificar a la hora de entrenar una red convolucional neuronal son el *padding* y el *stride*.

El *padding* se refiere a la cantidad de pixeles que se añaden artificialmente a una imagen cuando se procesa por el *kernel* de una convolución. Por ejemplo, la Ilustración 16 presenta un *padding* de 1, lo que significa que se añade filas y columnas de valor 1 alrededor de la imagen para permitir calcular el valor de la convolución en cada pixel de la imagen. Otra posibilidad muy utilizada es un *padding* de cero, entonces cada valor que se añade tendrá valor cero e impactara menos en teoría el resultado de la convolución obtenido. Estas dos posibilidades son ejemplos de *padding* constante, pero es posible tener un *padding* que no sea constante según los pixeles de los bordes de la imagen. Es el caso por ejemplo del *padding* por reflexión. Este método consiste en reflejar la fila (o columna) en el relleno que se crea alrededor de la imagen. Así se asegura que la transición sea suave entre la imagen original y el *padding* añadido. Este método suele mejorar los resultados del modelo obtenido pero añaden más complejidad a la etapa de *padding* que no se basa en un único valor. Esta complejidad junto con la del modelo puede alcanzar los limites computacionales del hardware que se utiliza. Añadir *padding* al entrenamiento de la red permite una reducción mínima del tamaño de la imagen en la capa de salida.

El *stride* corresponde al parámetro del filtro de la red neuronal que indica la cantidad de movimiento sobre la imagen. Por ejemplo, un valor del *stride* de uno indica que el filtro de convolución se mueve pixel a pixel (una unidad) cada ver. Si se incrementa el *stride*, se reduce el valor de la imagen

de salida obtenida. Este parámetro funciona en conjunto con el *padding*, el *stride* con valor alto reduce el tamaño de la imagen de salida cuando el *padding* tiene el efecto opuesto.

2.4. U-Net

Dentro de las redes neuronales convolucionales para la segmentación en imágenes médicas destaca el uso de algoritmos que emplean la arquitectura U-Net. La U-Net es una red neuronal desarrollada en **2015** basada en **convoluciones** y no utiliza una capa totalmente conectada [41].

Esta red se utiliza ampliamente para la segmentación de imágenes médicas. En comparación con otras redes existentes, requiere menos imágenes de entrenamiento y sus resultados de segmentación son más precisos. El objetivo de esta red es segmentar las imágenes utilizando el contexto, es decir, la información contenida en los píxeles vecinos del píxel de interés. La U-Net se descompone en **dos ramas**: una **rama contractiva** y una **rama extensiva** lo que le da su forma característica de la letra U como se ve en la Ilustración 21.

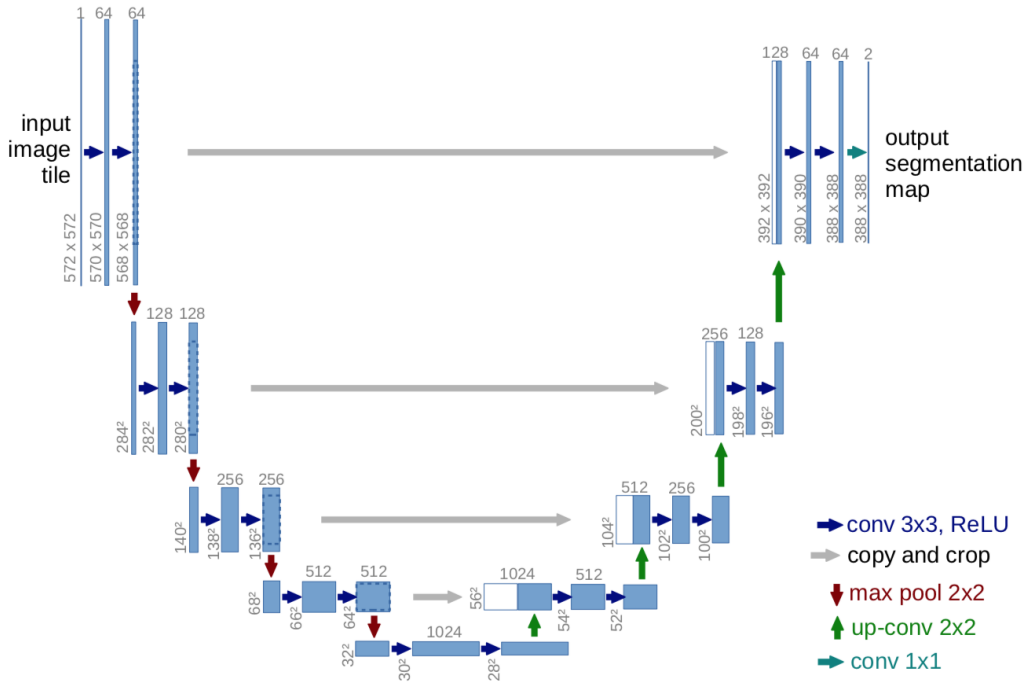


ILUSTRACIÓN 21: ESQUEMA DE LA U-NET DEL ARTICULO [41]

Estas dos ramas se subdividen en capas. Una capa de la rama de contracción consiste en una capa convolucional con un *kernel* de tamaño 3x3, cuya salida se somete a la función de activación ReLu (*Rectified Linear Unit*). Estas operaciones se repiten dos veces. Por último, se aplica un *maxpooling* de tamaño 2x2. Esta capa se repite un número definido de veces y luego los datos se redimensionan para que dejen de tener estructura matricial y pasen a ser un vector.

Una capa de la rama extensiva consiste en una deconvolución o convolución traspuesta empleando un *kernel* de tamaño 2x2. Esto duplica el número de características. Tener un número grande de características de esta manera permite que el conocimiento se propague a las capas más altas de la red. El resultado se concatena con la imagen situada en la misma fase de la rama de contracción. Las imágenes de la rama de contracción se reutilizan porque la operación de *max-pooling* sólo conserva la información de un píxel de cada cuatro, por lo que hay una pérdida de información.

Así pues, para construir una salida más precisa, se utiliza la información de la imagen original. Aunque este mecanismo permite llevar a cabo una mejor segmentación, requiere una capacidad grande de memoria porque el resultado de cada paso se mantiene en ella. Por lo tanto, es una desventaja si se utiliza una base de datos muy amplia y/o imágenes grandes.

Además, al concatenar es necesario recordar que hay que recortar la imagen de la rama de contracción porque cada convolución reduce el tamaño de la imagen debido al tratamiento de los bordes. De hecho, para esta red neuronal, no hay *zero padding* ni reflexión de los bordes en los espejos. Se elige recortar la imagen sin añadir ningún *padding* que permitiría conservar el tamaño inicial de la imagen.

Por último, hacemos dos convoluciones 3x3 cada uno, seguido de la aplicación de la función de activación ReLU.

En total, la arquitectura U-Net está compuesta por 23 capas convolucionales: 9 en la primera rama y 14 en la segunda. La segunda rama tiene más capas porque no sólo hay que volver a montar las capas de la rama de contratación, sino también los resultados de estas capas. Hay que tener cuidado con el tamaño de las imágenes de entrada para que se puedan realizar todas las operaciones, especialmente el max-pooling 2x2 que requiere una capa con dimensiones uniformes.

También se aplican deformaciones a las imágenes de entrada con un vector de desplazamiento aleatorio. Esto permite generar imágenes más variadas y de mayor tamaño, al tiempo que se limita el número de imágenes que hay que anotar inicialmente. La optimización de la red se basa en la función de pérdida de *cross entropy* que calcula la diferencia de valores entre la predicción y la estimación. El objetivo es minimizar esta función.

2.5. Bloques residuales

Para facilitar el entrenamiento de redes profundas como pueden ser las redes U-Net se puede usar una Residual U-Net como se propuso por primera vez en el artículo de extracción de carreteras [42]. Este algoritmo de *Deep learning* se va a denominar ResU-Net en este trabajo comparativamente con la U-Net presentada en la sección 2.4. Esta red se caracteriza por contar con unidades residuales que permiten facilitar la propagación del aprendizaje, permitiendo diseñar redes con menos parámetros y mejor rendimiento. Una unidad residual se diferencia de una unidad tradicional de una red neuronal (que solo alimenta a la siguiente neurona) por el hecho que una unidad residual alimenta a la siguiente y directamente a las capas que están a 2-3 saltos de distancia.

Estas capas residuales se pueden denominar *skip connections* en inglés para indicar que son conexiones que saltan algunas de las capas de la red neuronal para alimentar a las capas siguientes. Se introdujeron en la ResNet en 2015 para resolver el problema de clasificación de imágenes [43]. Es a partir de esta idea que se desarrolló la ResU-Net como arquitectura de red neuronal convolucional *encoder-decoder* a la cual se añaden capas residuales.

Cuando una arquitectura U-Net no es capaz de encontrar un mapeo más sencillo entre las entradas y las salidas, las capas residuales proponen una buena solución. La arquitectura de ResU-Net hace uso de conexiones de acceso directo para resolver el problema del desvanecimiento de gradiente.

En las unidades residuales, comparativamente con las unidades de doble convolución de la U-Net básica, se añade una función de mapeo identidad entre la entrada y la salida de la unidad como se ve en la Ilustración 22. La combinación de la U-Net con bloques residuales aporta conexiones entre los niveles bajos y altos de la red facilitando la propagación de la información sin degradación. Es una arquitectura de modelo que puede tener resultados incluso mejores que la U-Net clásica.

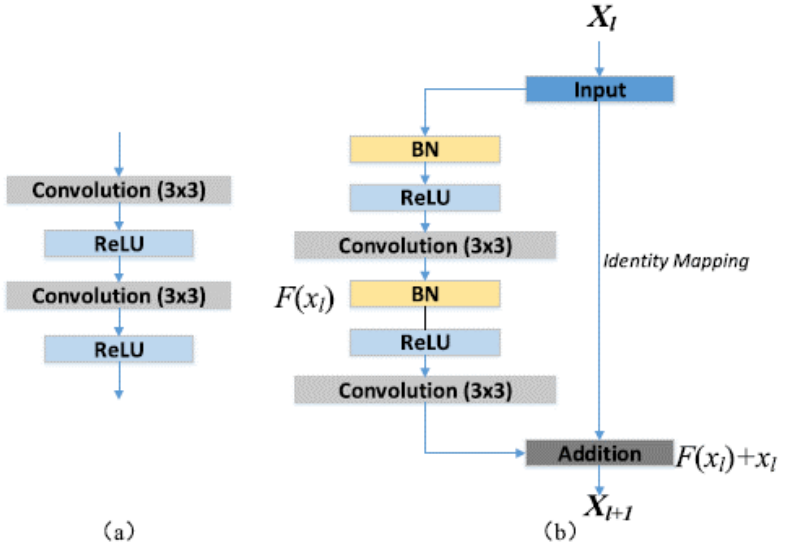


ILUSTRACIÓN 22: BLOQUES EN REDES NEURONALES (A) UNIDAD NEURONAL USADA EN LA U-NET (B) UNIDAD RESIDUAL PROPUESTA EN LA RESU-NET [42]

El modelo ResU-Net ya ha sido empleado en tareas de segmentación semántica para imágenes de histopatología como por ejemplo en el artículo [44] sobre su uso en cáncer de próstata y evaluación del score de Gleason. En este ejemplo, la ResU-Net permite obtener muy buenos resultados superando otras arquitecturas presentes en la literatura y proporcionando una localización detallada de los patrones.

3. Materiales

En esta parte se van a presentar las dos bases de datos utilizadas para este trabajo así que el preprocesado que se realiza. Se dispone de una base de datos anotadas que se denomina CLARIFYv1 en el trabajo siguiente y de una base de datos no anotada que se denomina CLARIFYv2. Se detalla también el entorno de programación que se elige para la implementación de las redes *encoder-decoder*. El dispositivo hardware utilizado esta descrito en esta parte también.

3.1. Presentación de la base de datos CLARIFYv1

Para el desarrollo de este trabajo, se ha utilizado la base de datos CLARIFYv1 que contiene 77 imágenes histopatológicas de biopsias de piel. Son imágenes obtenidas a partir de muestras de pacientes del servicio de Anatomía Patológica del Hospital Clínico Universitario de Valencia. Son muestras tomadas entre los años 1990 y 2017. Corresponden a 39 pacientes con lesiones melanocíticas spitzoides. De los 39 pacientes bajo estudio, 8 están diagnosticados con lesión melanocítica maligna (melanoma), 12 con lesión melanocítica benigna (nevus) y 19 con potencial maligno incierto (PMI). Al tener biopsias de tumores sanas y otras patológicas, la base de datos considerada presenta una variabilidad importante. Al momento de tratar las imágenes se han anonimizado previamente y se desconoce cuáles corresponden a biopsias de melanoma y cuáles no.

Estas imágenes de diferentes zonas del cuerpo presentan una tinción mediante hematoxilina eosina. La intensidad de la tinción varia de una imagen a otra según el tipo de tejido, el operador que ha preparado las tinciones y el microscopio que sirve a la digitalización de las imágenes. También el tamaño de las imágenes varía de un paciente a otro y según el número de cortes en una lámina. Las WSI de la base de datos son imágenes de alta resolución difíciles de procesar por modelos de *Deep Learning* por su tamaño, por lo que se suelen dividir en *patches* para su estudio.

Las imágenes histopatológicas con las que se trabaja en este proyecto se encuentran digitalizadas a una magnificación de 40x con el objetivo de disponer de una resolución espacial lo más alta posible y captar así los detalles más finos de las imágenes.

Un patólogo anota estas imágenes, indicando qué zona corresponde a la epidermis lo que permite obtener la máscara de epidermis real o *ground truth* en inglés. Para eso usa la aplicación *MicroDraw* [45] desarrollada sobre la librería *OpenSeadragon* que permite una visualización multi-resolución de las imágenes histológicas sin pérdida de calidad. La plataforma *MicroDraw* está escrita en JavaScript, empleando HTML5, CSS3 y jQuery.

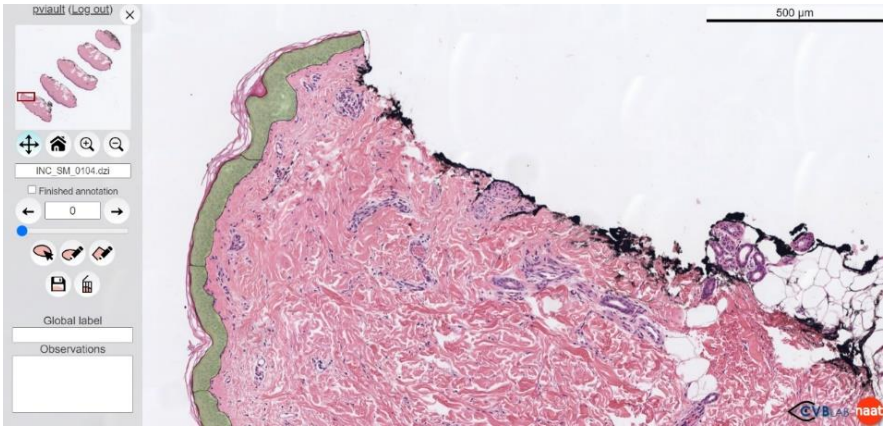


ILUSTRACIÓN 23: EJEMPLO DE PÁGINA DE ANOTACIÓN CON *MICRODRAW* (ILUSTRACIÓN PROPIA)

Debido a las limitaciones de hardware es necesario realizar un ajuste en el tamaño de las imágenes. Este cambio es hecho sin recortar las imágenes para evitar la pérdida de información. Se cambia el tamaño de las imágenes a 512*512 pixeles. Pero las imágenes se distorsionan, se pierde información y precisión. Estas imágenes redimensionadas se van a llamar imágenes completas en la memoria en el sentido que son las WSIs vistas desde mayor distancia y con un menor nivel de detalle.

A continuación, para realizar el entrenamiento de la red se divide de manera aleatoria la base de datos. Así se obtienen una partición de entrenamiento con el 70% de las imágenes de la base de datos, validación con un 15% de las imágenes y test el 15% restante. Se realiza la partición con imágenes que corresponden a diferentes pacientes para evitar el sesgo de mezclar muestras del mismo paciente en los diferentes conjuntos de datos. La repartición realizada esta ilustrada en la Ilustración 24, hay 45 imágenes de entrenamiento, 10 imágenes de validación y 11 imágenes de test.

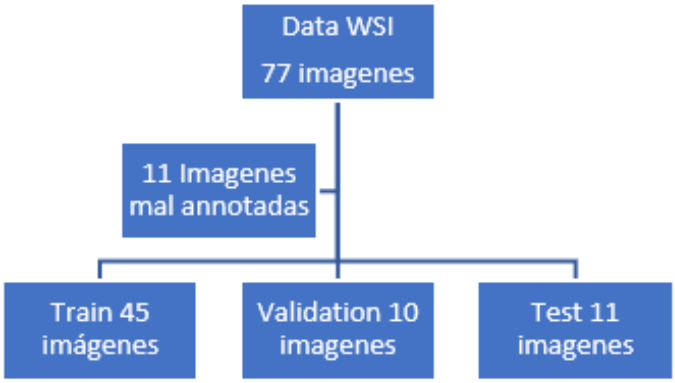


ILUSTRACIÓN 24: PARTICIÓN DE LA BASE DE DATOS CLARIFYV1

Como se observa en la Tabla 2, los pixeles de la epidermis representan un 10% aproximadamente de la WSI y esta proporción se conserva bien entre los diferentes conjuntos de datos. Por tanto, no existe diferencia estadísticamente significativa entre los subgrupos de imágenes, lo que asegura la buena repartición de la información de interés entre los diferentes conjuntos.

TABLA 2: PROPORCIÓN DE PIXELES QUE PERTENECEN A LA EPIDERMIS EN LAS IMÁGENES DE LA BASE DE DATOS CLARIFYV1

	Entrenamiento	**Validación**	**Test**
Porcentaje de pixeles de la epidermis	11.98%	8.04%	9.14%

En la Ilustración 25 se aprecia la repartición de la epidermis en una muestra, de esta imagen 40% corresponden al fondo, 47% al tejido de la biopsia (dermis e hipodermis) y 13% a la epidermis. De una imagen a otra la proporción de epidermis no es exactamente la misma, pero destaca que es una estructura siempre presente en menor proporción que el resto de los tejidos en las imágenes.

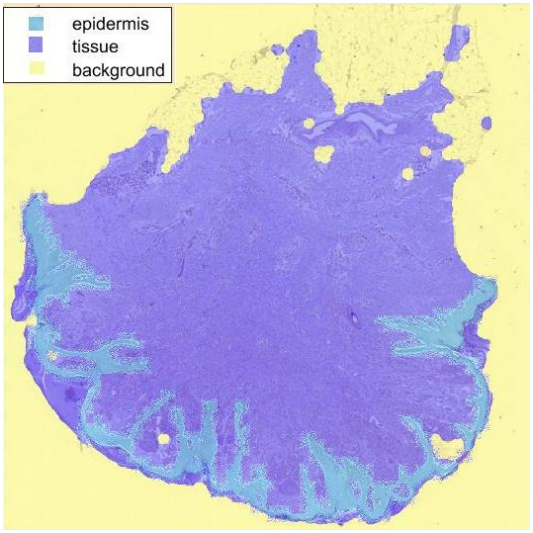


ILUSTRACIÓN 25: EJEMPLO DE LA REPARTICIÓN DE LAS DIFERENTES ESTRUCTURAS EN LAS IMÁGENES (ILUSTRACIÓN PROPIA)

En la base de datos empleada hay mucha diversidad de tamaño de las imágenes, de media son de tamaño 5622*4646 pero la más pequeña es de ancho 512. El cambio de resolución de las imágenes a 512*512 introduce distorsión y pérdida de la fidelidad de las imágenes de entrada como se puede ver en la Ilustración 26.

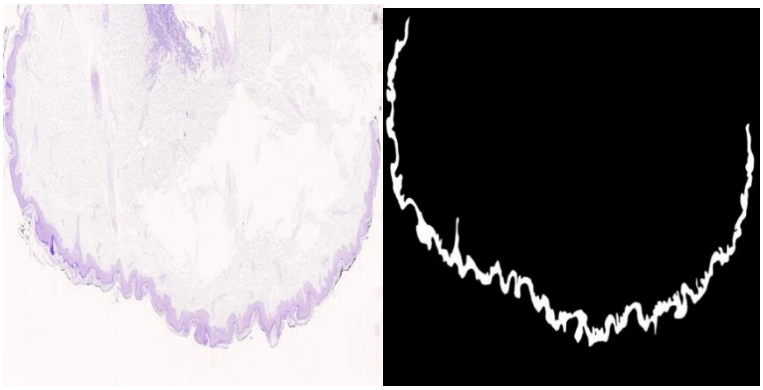


ILUSTRACIÓN 26: EJEMPLO DE DISTORSIÓN DE UNA IMAGEN DE ENTRADA (ILUSTRACIÓN PROPIA)

Se tendrá únicamente una segmentación gruesa de la epidermis, pero resulta importante para indicar en que zonas de la imagen hay que centrarse para la segmentación de precisión. Permite quitar del análisis una grande parte de la imagen y eso reduce el tiempo de computación de la inferencia.

Como la intensidad y los colores de la tinción varían mucho de una imagen a otra, como se ve en la Ilustración 27, se plantea llevar a cabo una normalización de la tinción a una imagen de referencia. Se ha probado hacer con el módulo *torchstain*.

Este algoritmo como indicado en el artículo [46], se basa en la conversión de las imágenes RGB a densidad óptica (OD en inglés). Suponen que existe un vector específico que representa cada una de las tinciones (hematoxilina y eosina) y que el color de cada pixel en el espacio OD es una combinación lineal de los dos vectores. Usan versiones robustas de los mínimos y máximos de la distribución de pixeles para determinar estos vectores. Antes utilizan un umbral para quitar los pixeles de poca inversión (con OD cerca de cero), esto contribuye a una mejor estabilidad del algoritmo. Para obtener los dos vectores se forma un plano a partir de los dos vectores correspondientes a los dos mayores valores singulares de la descomposición SVD de los píxeles transformados en OD. Mediante esta técnica obtienen los vectores de la tinción que les permite convertir una imagen a su tinción

normalizada. Para eso, utilizan el camino más corto entre los dos vectores. La nueva imagen con tinción normalizada se obtiene proyectando los pixeles sobre este camino.

En este artículo, demuestran que aplicando la técnica propuesta de normalización de la tinción entre las diferentes imágenes logran a discriminar imágenes de melanoma de imágenes de nevus. Lo logran gracias al método *Distance Weighted Discrimination*. Además, la normalización automática de la tinción permite un nivel consecuente de reproducibilidad comparativamente con métodos de selección manual.

Pero este algoritmo necesita una única imagen de referencia. Este proceso introduce mucho sesgo al elegir una única imagen que debe representar la tinción requerida. En el caso de la base de datos previamente presentada, el uso de la normalización de la tinción resulta en una pérdida importante de contraste como se ve en la ilustración 26. Por eso y el sesgo que puede introducir se ha finalmente elegido de no usar una normalización de la tinción.

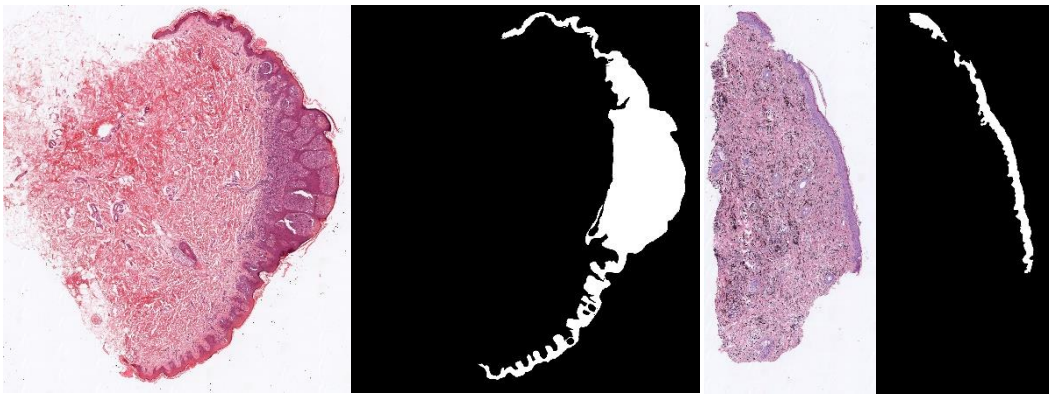


ILUSTRACIÓN 27: EJEMPLOS DE TINCIÓN DIFERENTE Y TAMAÑO DE IMAGEN DIFERENTE (ILUSTRACIÓN PROPIA)

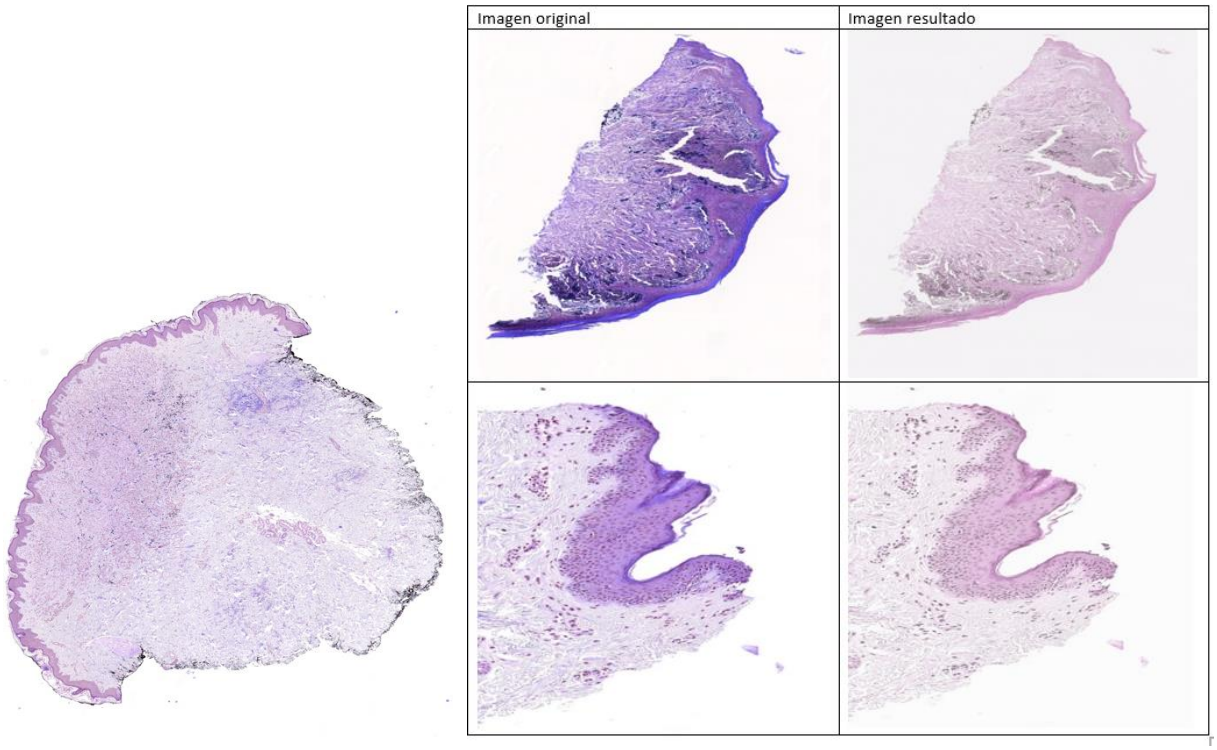


ILUSTRACIÓN 28: EJEMPLO DE NORMALIZACIÓN DE LA TINCIÓN CON EL MÓDULO *TORCHSTAIN* (A LA IZQUIERDA LA IMAGEN DE REFERENCIA Y A LA DERECHA TABLA DE IMÁGENES ORIGINALES EN LA PRIMERA COLUMNA E IMÁGENES NORMALIZADAS EN LA SEGUNDA COLUMNA) (ILUSTRACIÓN PROPIA)

Las 8 imágenes que tienen un problema de anotación son como la Ilustración 29. Estas imágenes entonces solo se van a usar en inferencia**,** pero no en el conjunto de entrenamiento.

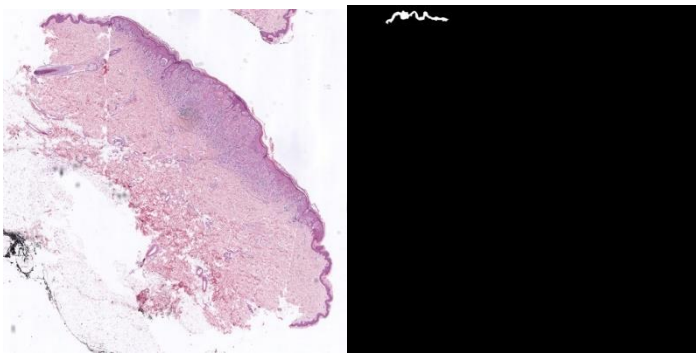


ILUSTRACIÓN 29: IMAGEN CON PROBLEMA DE ANOTACIÓN (ILUSTRACIÓN PROPIA)

Y también hay ciertas imágenes que no parecen relevantes para la tarea de segmentación de la epidermis como se ven en la Ilustración 30.

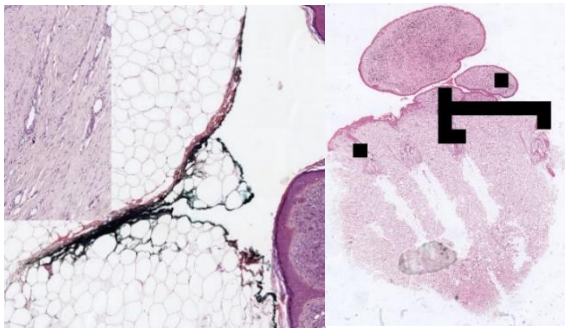


ILUSTRACIÓN 30: IMÁGENES CON ANOMALÍAS (ILUSTRACIÓN PROPIA)

Las imágenes se dividen en *patches* para permitir la segmentación precisa de la epidermis. Los *patches* se obtienen solo en las regiones que tienen epidermis en la máscara de *ground truth* y con un solapamiento del 50%. El estudio se centra en los *patches* con anotaciones. Esto permite ejecutar el algoritmo con mayor resolución, pero solo en zonas previamente seleccionadas como posibles candidatas de contener epidermis. Es muy importante este paso ya que si no el volumen de *patches* implicaría un gran coste computacional.

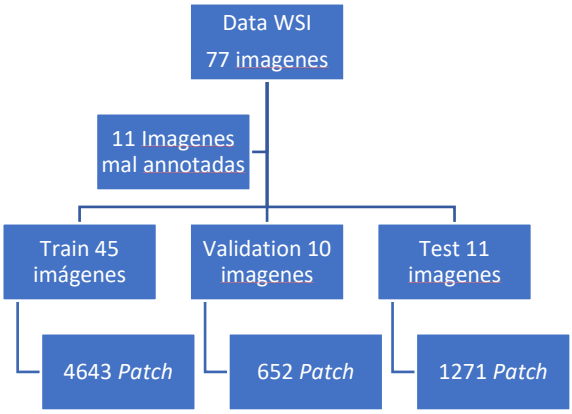


ILUSTRACIÓN 31: PARTICIÓN DE LOS DATOS A NIVEL DE PATCH DE LA BASE DE DATOS CLARIFYV1

La partición de los datos sigue el mismo proceso aleatorio que detallado previamente con un 70% de datos de entrenamiento, un 15% para los datos de validación y un 15% para los datos de test.

3.2. Presentación de la base de datos CLARIFYv2

La base de datos CLARIFYv2 contiene 82 imágenes y fue anotada por el estudiante bajo la supervisión del patólogo con la herramienta *MicroDraw*. Esta base de datos es similar a CLARIFYv1, también son imágenes histopatológicas de biopsias de piel del hospital clínico universitario de Valencia. Son imágenes provenientes de pacientes con las mismas patologías que la base de datos CLARIFYv1 conteniendo biopsias de melanoma spitzoide. Estas imágenes que no estaban anotadas antes de este trabajo de fin de master y por lo tanto, son de especial interés para servir de base de datos de validación del modelo implementado de segmentación automática de la epidermis. La Ilustración 32 muestra un ejemplo de las imágenes que se trata de anotar.

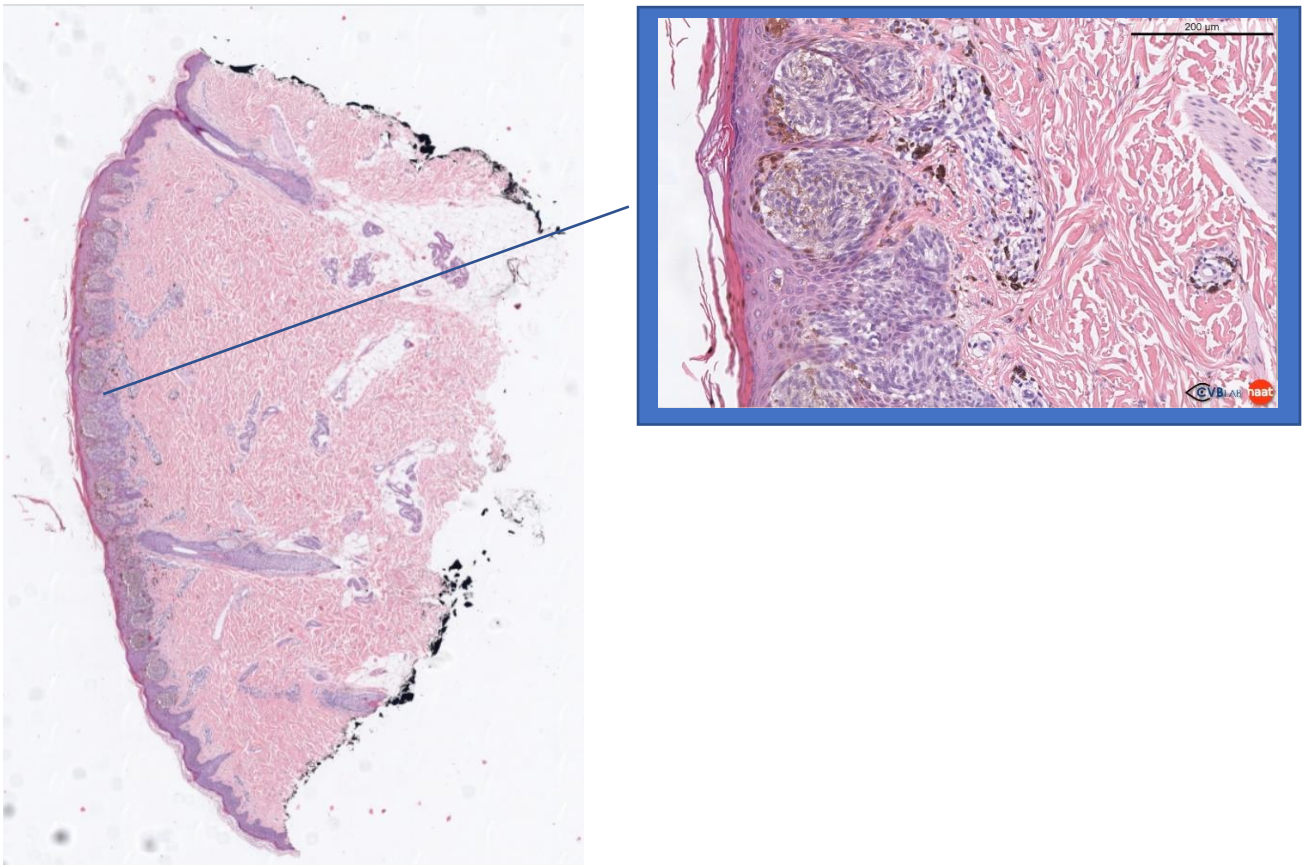


ILUSTRACIÓN 32: EJEMPLO DE UNA IMAGEN CON ZOOM EN LA ZONA DE LA EPIDERMIS QUE SE TRATA DE DETECTAR (ILUSTRACIÓN PROPIA)

La epidermis se nota con su color más oscuro pero las lesiones melanocíticas epidérmicas o compuestas hacen más compleja la tarea de segmentación manual de la epidermis. La dificultad de distinguir el límite entre la epidermis y la dermis viene en parte de la presencia de crestas epidérmicas y de papilas dérmicas como explicado en la sección 1.2.2.

En esta nueva base de datos utilizada, las WSI anotadas presentan diferentes regiones que cada una corresponde a la biopsia de piel de una zona de la lesión del paciente, como lo muestra la Ilustración 33 que presenta 5 regiones dentro de la WSI. Entonces a partir de las 82 imágenes, destacan 245 nuevas regiones de biopsias de piel eliminando las regiones repetidas o las que tenían anomalías en la tinción por ejemplo que no se han podido anotar.

ILUSTRACIÓN 33: IMAGEN 0 DE CLARIFYV2 (ILUSTRACIÓN PROPIA)

Al igual que para la base de datos CLARIFYv1, se realiza una partición aleatoria de los datos para establecer los conjuntos de datos de entrenamiento, validación y test. Las mismas proporciones son utilizadas, 70%, 15% y 15% respectivamente. El número de regiones obtenidas en cada grupo se detalla en la Ilustración 34. Esta base de datos es importante para validar, entrenar y mejorar el modelo de segmentación automática de la epidermis en imágenes con lesiones melanocíticas spitzoides porque presenta un número mayor de muestras y entonces más variabilidad de los datos.

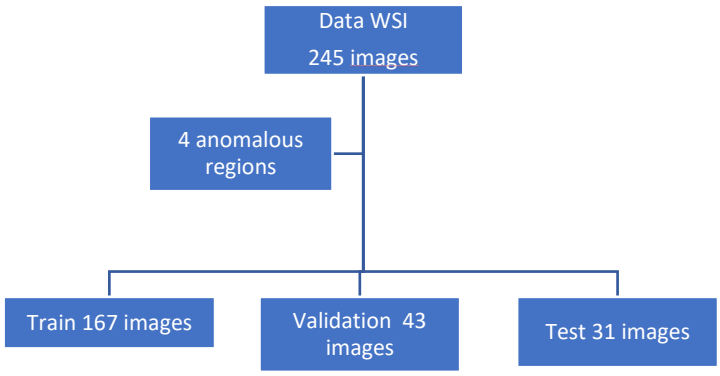


ILUSTRACIÓN 34: PARTICIÓN DE LA BASE DE DATOS CLARIFYV2

Se realiza una partición similar para los *patches* obtenidos. El número de *patches* de cada base de datos y de cada conjunto se presenta en la Tabla 3. Se ve que la base de datos CLARIFYv2 es mucho más amplia que CLARIFYv1 para tener conjuntos de imágenes que representan mejor la variabilidad de la realidad clínica de las diferentes lesiones melanocíticas benignas y malignas en las biopsias de piel.

TABLA 3: NÚMERO Y REPARTICIÓN DE LOS PATCHES DE CADA BASE DE DATOS ENTRE LOS CONJUNTOS DE ENTRENAMIENTO, VALIDACIÓN Y TEST

BASE DE DATOS	ENTRENAMIENTO	VALIDACIÓN	TEST
CLARIFYV1	4643	652	1271
CLARIFYV2	16 149	3 245	3 748

A continuación se presenta la partición obtenida para las dos bases de datos presentadas para la técnica de validación cruzada explicada en la sección 4.7.

Para la implementación de la técnica de la validación cruzada en el entrenamiento, se dividen aleatoriamente las dos bases de datos en 5 grupos conservando imágenes para el conjunto de test, lo que da la repartición mostrada en la Tabla 4. Entonces para el conjunto de entrenamiento se utiliza 44 imágenes de la base de datos CLARIFYv1, respectivamente 36 de la base de datos CLARIFYv2; 11 imágenes de validación para la base de datos CLARIFYv1, respectivamente 9 para la base de datos CLARIFYv2 y 11 imágenes de test para la base de datos CLARIFYv1, respectivamente 9 imágenes de test para la base de datos CLARIFYv2.

TABLA 4: REPARTICIÓN DE LAS IMÁGENES EN GRUPOS

BASE DE DATOS	CLARIFYV1	CLARIFYV2
GRUPO 1	11	10
GRUPO 2	11	9
GRUPO 3	11	9
GRUPO 4	11	9
GRUPO 5	11	9
TEST	11	9

3.3. Presentación del entorno de programación utilizado

Para la implementación del modelo se ha utilizado el lenguaje de programación Python que propone una gran variedad de librería *open source* para el análisis predictivo de datos, entre otras muchas funcionalidades. Python es un lenguaje de programación interpretado donde la sintaxis, claramente separada de los mecanismos de bajo nivel, permite una fácil introducción a los conceptos básicos de programación. También permite desarrollar modelos de inteligencia artificial usando Tensorflow y Pytorch como se presenta a continuación.

Las librerías utilizadas en este trabajo son:

- **Cv2**, **PIL** y **scikit-image** como librerías de procesado de imágenes
- **Numpy** como librería para manipular matrices multidimensionales y funciones matemáticas que operan sobre estas matrices
- **Matplotlib.pyplot** para la creación de imágenes y gráficas y su visualización
- **Os** para usar funciones del sistema operativo, sobre todo para la manipulación de ficheros y carpetas
- **Scikit-learn** y **torch** para el aprendizaje automático
- **Torchsummary** para visualizar la arquitectura del modelo implementado
- **Albumentations** como herramienta de visión por ordenador que potencia el rendimiento de las redes neuronales convolucionales profundas
- **Tqdm** para barra de progresión

- **Torchstain** como herramientas de normalización compatibles con Pytorch para imágenes histopatológicas [46]

Como entorno de desarrollo integrado para programar en Python se usó PyCharm. Este software permite el análisis del código y contiene un depurador gráfico.

La arquitectura de la red neuronal profunda utilizada fue implementada en **Pytorch**. PyTorch es una biblioteca de aprendizaje automático de código abierto en Python basada en Torch, desarrollada por Facebook. PyTorch se utiliza para realizar los cálculos tensoriales necesarios para el aprendizaje profundo.

Además, como librería para el aprendizaje profundo, se ha hecho uso de la librería NVIDIA Cuda® Deep Neural Network (cuDNN). Es una biblioteca que permite acelerar la ejecución del entrenamiento de redes neuronales profundas en la GPU. Una GPU (unidad de procesamiento grafico en inglés) es una unidad de computación que realiza las funciones de procesamiento de imágenes, mostrarlas en la pantalla o escribirlas en el almacenamiento masivo.

3.4. Hardware utilizado

Este trabajo se realiza gracias a un servidor de computación que se conecta a la NAS mediante una conexión SSH y con Docker. **Docker** es un proyecto de código abierto que automatiza el despliegue de aplicaciones dentro de contenedores de software, proporcionando una capa de virtualización de aplicaciones. La Ilustración 35 representa esta conexiones. Se usa el software MobaXterm para establecer estas conexiones.

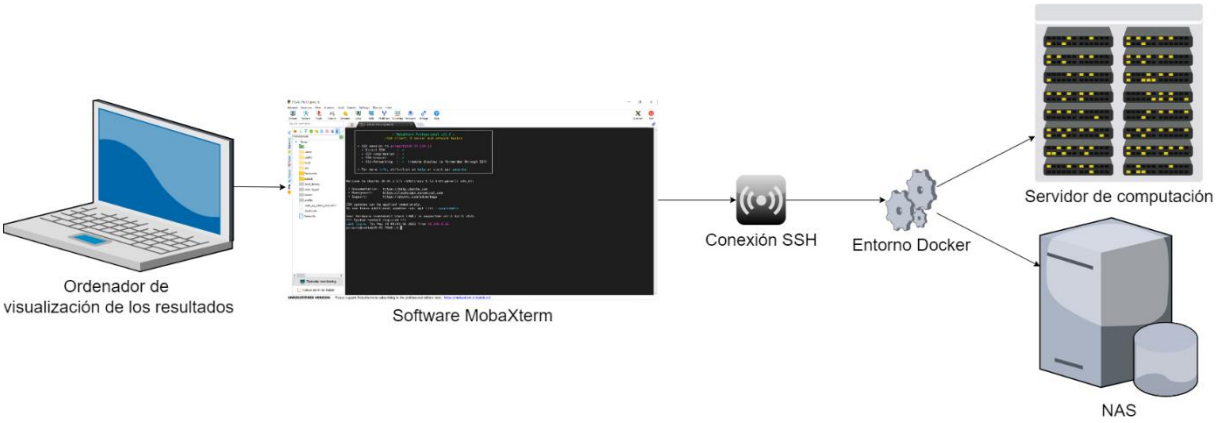


ILUSTRACIÓN 35: ESQUEMA DEL HARDWARE UTILIZADO (ILUSTRACIÓN PROPIA)

Las características del ordenador que ha servido a la programación del modelo, al preprocesado y post procesado de las bases de datos tal y como a la visualización de los resultados se presentan en la Tabla 5 junto con las características del servidore de computación.

TABLA 5: CARACTERÍSTICAS DEL ORDENADOR Y DEL SERVIDOR DE COMPUTACIÓN

	Procesador	**Memoria caché**	**Frecuencia**	**Sistema operativo**	**Memoria RAM**
Ordenador de visualización de los resultados y de programación	Intel Core i7-1165G7	12 MB	4,7GHz	Windows 10 de 64 bits	16GB
Servidor de computación	Intel i7	/	4.20GHz	Linux	32GB

En cambio, para el entrenamiento de redes neuronales profundas sobre imágenes grandes y de alta resolución se necesita servidores con unidades de procesamiento gráficos más potentes (GPU). Así, se ha utilizado para llevar a cabo esta etapa del trabajo un servidor de computación de altas prestaciones perteneciente al grupo de investigación del CVBLab. Este servidor de computación tiene 2 GPU y tarjetas gráficas NVIDIA Titan XP.

En paralelo se necesita también un servidor de almacenamiento de las imágenes porque el servidor previamente presentado solo se usa como recurso de computación pero no cuenta con los datos presentados. Estas imágenes ocupan mucho espacio en disco y por eso se ha usado la NAS Synology DS918 que tiene las capacidades suficientes para almacenarlas. NAS, *network-attached storage* en inglés, es un dispositivo de almacenamiento conectado a una red que permite almacenar y recuperar los datos en un punto centralizado para usuarios autorizados de la red.

4 Metodología

4.1. *Framework* propuesto para la segmentación de la epidermis en imágenes histológicas

En este trabajo se propone un enfoque a dos niveles de resolución para realizar la segmentación precisa de la epidermis. En el primer nivel se usa toda la imagen haciéndole un redimensionamiento a 512*512 para que tenga un tamaño asumible por las limitaciones de hardware. Las imágenes de entrada son normalizada a valores entre 0 y 1. Si se usan las imágenes sin cambiarla, los cálculos de los valores numéricos altos pueden resultar complejo y ocupar mucho espacio en la memoria. Para reducir este fenómeno, esta normalización permite que los números sean pequeños y el cálculo se hace de forma más fácil y rápida.

Con esta primera red se obtiene una segmentación gruesa de la epidermis. A partir de esta segmentación se sacan *patches* de 512*512 dentro de la zona segmentada como epidermis con un solapamiento del 50%. Luego los *patches* entran en la segunda red de precisión y la reconstrucción de las máscaras obtenidas permite obtener la segmentación precisa de la epidermis deseada. Este proceso se detalla también en la Ilustración 36.

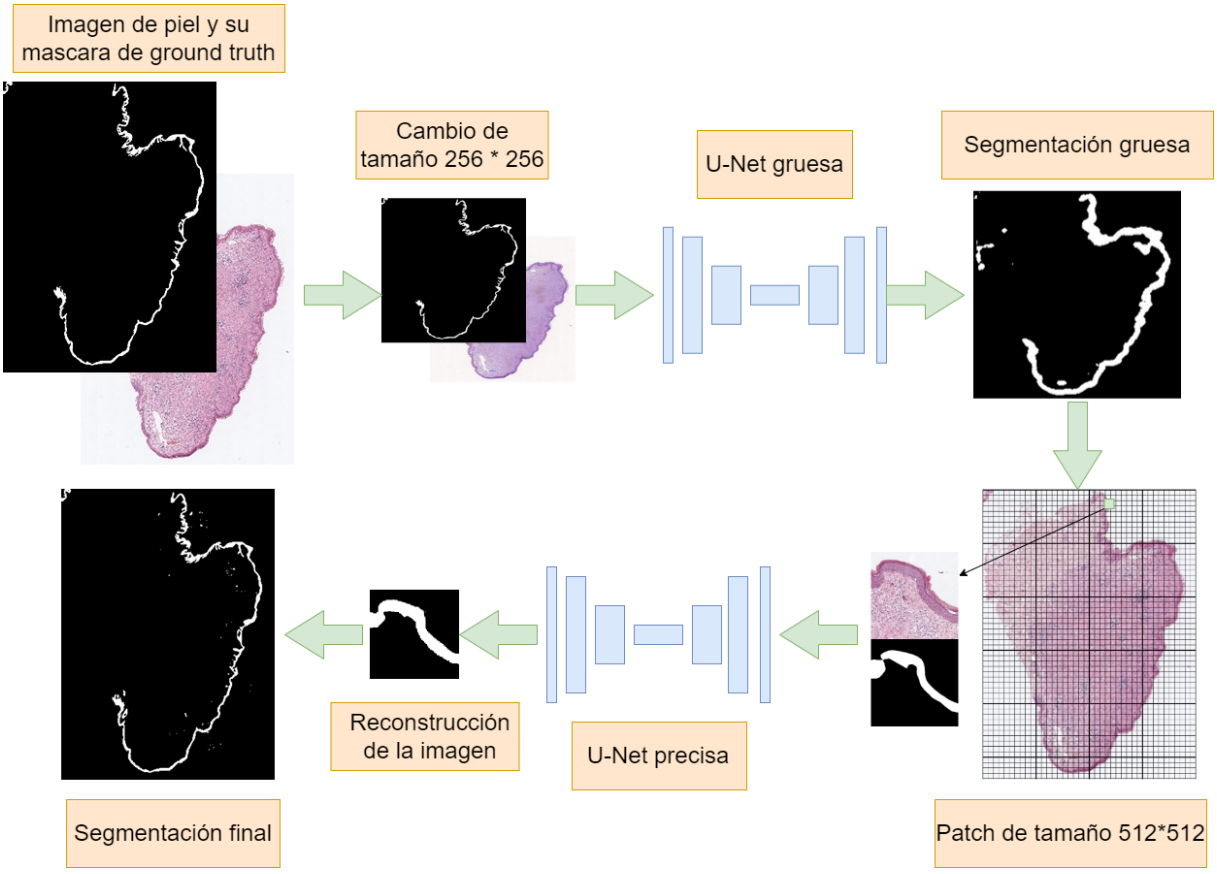


ILUSTRACIÓN 36: ESQUEMA DEL PROCESO SEGUIDO EN ESTE TRABAJO

Cada red neuronal utilizada ha sido optimizada con hiperparámetros que permitan un entrenamiento riguroso pero también en un tiempo de computación óptimo. Estos hiperparámetros se detallan en la Tabla 6. Se precisa para cada experimento realizado el nivel de resolución, gruesa siendo con la imagen completa para tener una idea de donde está la epidermis y precisa los modelos

a nivel de *patch* la base de datos que sirve para entrenar el modelo, si la validación cruzada fue utilizada, el número de épocas, el *learning rate*, el *batch size* y el tamaño de las imágenes en entrada de la red convolucional *encoder-decoder* con el tipo de arquitectura. La base de datos CLARIFYv1 es simplificada en base de datos 1 y la base de datos CLARIFYv2 de la misma manera en base de datos 2.

TABLA 6: HIPERPARÁMETROS DE LOS DIFERENTES EXPERIMENTOS REALIZADOS

Nivel de resolución	Experimento	Tipo de arquitectura de red	Base de datos	Validación cruzada	Numero épocas	Learning rate	Batch size	Tamaño imagen
Segmentación gruesa	1	U-Net	1	No	150	0,0001	16	256
	2	Residual U-Net	1	No	150	0,0001	16	256
	3	U-Net	1	Yes	150	0,0001	16	256
	4	Residual U-Net	1	Yes	150	0,0001	16	256
	5	U-Net	2	Yes	150	0,0001	16	256
	6	Residual U-Net	2	Yes	150	0,0001	16	256
	7	U-Net	1&2	Yes	150	0,0001	16	256
	8	Residual U-Net	1&2	Yes	150	0,0001	16	256
Segmentación precisa	9	U-Net	1	No	50	0,0001	4	512
	10	Residual U-Net	1	No	50	0,0001	4	512
	11	U-Net	2	No	50	0,0001	4	512
	12	Residual U-Net	2	No	50	0,0001	4	512
	13	U-Net	1&2	No	50	0,0001	4	512
	14	Residual U-Net	1&2	No	50	0,0001	4	512

Como primer enfoque se hace un entrenamiento clásico dividiendo entre datos de entrenamiento, validación y test. A continuación, para una mayor precisión de los resultados obtenidos se hacen los mismos experimentos, pero utilizando la técnica de validación cruzada como se explica en la sección 4.7 de este documento. Los experimentos 5 a 8 y 11 a 14 utilizan como datos de entrenamiento imágenes de la base de datos CLARIFYv2 comentada en la parte 3.2.

4.2. Arquitectura de la red y tipo de entradas

El entrenamiento se realiza con una red neuronal convolucional con arquitectura **U-Net** para la segmentación de imágenes. Las imágenes de entrada son de tamaño **256*256** como redimensionamiento de la WSI y se usa un ***batch size*** **de 16**. Para los entrenamientos a nivel de *patch*, las imágenes de entrada son de tamaño **512*512** con un ***batch size*** **de 4**.

En el caso de este proyecto, ya que se trata de una red *encoder-decoder*, el modelo debe reconstruir la imagen de salida al tamaño que tiene en entrada y para eso guarda en memoria la información de cada convolución de la parte *encoder*. Eso hace que la GPU utilizada con una arquitectura de red convolucional U-Net y ResU-Net solo permite un *batch size* de 4 para no superar el límite de memoria. Si la imagen de entrada es de 256*256, 16 imágenes caben en un batch pero si el tamaño es de 512*512, este número se reduce a 4. Un valor más grande podría permitir un entrenamiento más rápido, robusto y preciso pero con los limites computacionales de los servidores no es posible.

Se calculan todos los pesos de la red partiendo de pesos aleatorios, no se carga ningún tipo de red preentrenada sino que todos los pesos se ajustan con el entrenamiento realizado.

Se define el bloque de base de la arquitectura de la red neuronal *encoder-decoder* utilizada como una doble convolución que tiene la forma siguiente:

- Una convolución en 2D con un *kernel* de tamaño 3*3, un *stride* de 1 para la convolución y un *padding* de 1. Se añade un sesgo posible de entrenar a la salida. Se utiliza la función Conv2d del módulo torch.nn que realiza la operación siguiente.

  Si se nota (N, C_{in}, H, W) el tamaño de la entrada y $(N, C_{out}, H_{out}, W_{out})$ el tamaño de la salida, la operación realizada se puede escribir así:

$$\boldsymbol{out}(\boldsymbol{N_i}, \boldsymbol{C_{out_j}}) = \boldsymbol{bias}(\boldsymbol{C_{out_j}}) + \sum_{\boldsymbol{k=0}}^{\boldsymbol{C_{in}-1}} \boldsymbol{weight}\left(\boldsymbol{C_{out_j}}, \boldsymbol{k}\right) * \boldsymbol{input}(\boldsymbol{N_i}, \boldsymbol{k})$$

ECUACIÓN 3: CONVOLUCIÓN

  Donde

  - $*$ es el operador de correlación cruzada 2D,
  - N es un tamaño de lote,
  - C_{in} y C_{out}denotan un número de canales,
  - H es una altura de planos de entrada en píxeles,
  - W es la anchura en píxeles.

- Un *batch normalization* en 2D que permite paliar al reto de la distribución de las entradas de cada capa cambia durante el entrenamiento al cambiar los parámetros de las capas anteriores. Por eso se necesita una normalización de las entradas de las capas para cada *minibatch* de entrenamiento. El *batch normalization* permite usar un *learning rate* más alto y sirve de regularizador como explicado en el artículo [47]. Se realiza el cálculo siguiente restando la media y dividiendo por la desviación estándar regularizada junto a un factor épsilon sobre los *mini-batches* para prevenir errores si la varianza tiene valor cero. Gamma y Beta son parámetros que se aprenden como vectores. Per defecto, gamma se inicializa a uno y beta a cero.

$$\boldsymbol{y} = \frac{\boldsymbol{x} - \boldsymbol{E}[\boldsymbol{x}]}{\sqrt{\boldsymbol{Var}[\boldsymbol{x}] + \boldsymbol{\epsilon}}} * \boldsymbol{\gamma} + \boldsymbol{\beta}$$

ECUACIÓN 4: BATCH NORMALIZATION

- Activación con la función ReLU previamente presentada.

Y este proceso se repite dos veces por eso se considera como una doble convolución.

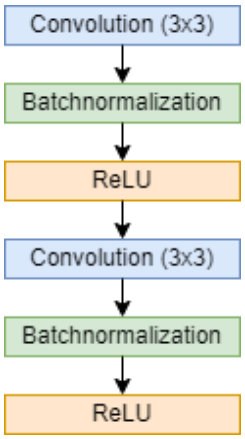


ILUSTRACIÓN 37: ESQUEMA DE LA DOBLE CONVOLUCIÓN UTILIZADA (ILUSTRACIÓN PROPIA)

4.3. ResU-Net

Como se ha expuesto anteriormente en el marco teórico, en este trabajo se implementa la arquitectura de ResU-net para comprobar si mejora los resultados de una red U-Net clásica. Consiste en usar la misma arquitectura global de red *encoder-decoder*, pero cambiar los bloques de convolución para añadir conexiones residuales como se muestra en la Ilustración 38.

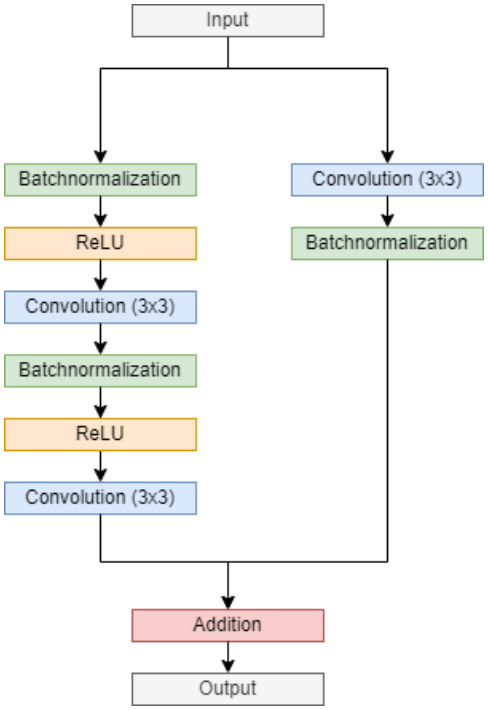


ILUSTRACIÓN 38: ESQUEMA DEL BLOQUE RESIDUAL UTILIZADO (ILUSTRACIÓN PROPIA)

El entrenamiento de la ResU-Net se realiza de igual manera que el de la U-Net con los mismos tipos de entrada, la misma función de activación, el mismo optimizador, el mismo *learning rate*, el método de *one hot encoding* y la técnica de validación cruzada como se presenta a continuación.

4.4. Métrica del entrenamiento y función de pérdida

La métrica elegida es el coeficiente ***dice*** y 1 – *dice* es la función de pérdidas del modelo.

El índice de Sørensen-Dice es un indicador estadístico que mide la similitud de dos muestras. El índice *dice* es 2 veces la intersección entre las 2 imágenes binarias dividido por la unión de estas imágenes que aquí es la suma de los pixeles de cada imagen. Es una métrica comúnmente usada y recomendada en tareas de segmentación con imágenes médicas juntos con el Jaccard Index [48].

$$Dice = \frac{2.|X \cap Y|}{|X| + |Y|}$$

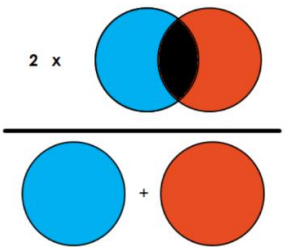


ILUSTRACIÓN 39: ESQUEMA QUE REPRESENTA COMO CALCULAR EL COEFICIENTE DE *DICE* [49]

La función de pérdida usada es un *dice* ponderado. Se corresponde con la métrica del índice de *dice* ya presentado, pero tiene en cuenta el desbalance de clase de la base de datos. Existen más pixeles que pertenecen al fondo que a la epidermis, lo que genera un desequilibrio en las proporciones de las dos clases consideradas. Este cambio se produce en la manera de calcular el *dice* en Pytorch que realiza la suma según los canales que corresponden para paliar al desbalance de clases.

4.5. Hiperparámetros del entrenamiento

El método de optimización elegido es **Adam** (Adaptative Moment Estimation) con un *learning rate* de **1*10^(-4)**. Este *learning rate* ha sido elegido empíricamente haciendo varios ensayos y viendo que corresponde a un valor de referencia en la literatura que permite que el modelo aprenda sobre los datos de entrenamiento de manera lo suficiente rápida para no tener un entrenamiento de larga duración. No obstante, este valor no es demasiado grande para que el algoritmo de optimización caiga en óptimo local.

El método de optimización Adam es una técnica de optimización basada en el algoritmo Adam que conta con dos componentes principales, un componente de momentum y un componente de *learning rate* adaptativo como descrito en el artículo [50].

El algoritmo de optimización Adam proviene del algoritmo de descenso de gradiente estocástico. El descenso de gradiente permite resolver la optimización de una función parametrizada que se trata de minimizar o maximizar. Este método es eficaz sobre todo en los casos en los cuales la función objetivo es diferenciable con respecto a sus parámetros.

Sin embargo, esta técnica puede alcanzar los límites de su eficiencia en casos con ruido y espacios de parámetros de alta dimensión. El método Adam permite la optimización eficiente usando los gradientes de primer orden y pocos requisitos de memoria. Se basa en el cálculo de *learning rate* adaptativos a partir de las estimaciones de los primeros y segundos momentos de los gradientes mientras que el método de descenso de gradiente mantiene un solo *learning rate*. El algoritmo de Adam presenta muchas ventajas como la de funcionar con gradientes dispersos.

En el caso de este trabajo con redes neuronales convolucionales, el reparto de pesos en capas que no son totalmente conectadas da lugar a gradientes muy diferentes en las distintas capas. El método de Adam es especialmente eficaz en el caso de red neuronal convolucional profunda permitiendo un entrenamiento más rápido que usando el descenso de gradiente como método de optimización de los parámetros. Entonces es un método que conviene de usar en casos de aprendizaje automático con espacios de parámetros de alta dimensión como sucede en este proyecto.

4.6. Codificación *one hot encoding*

Para mejorar los resultados de la segmentación con redes neuronales convolucionales se puede elegir predecir dos etiquetas de salida y poner en entrada al modelo dos clases como etiqueta. Para eso se pone en entrada un canal correspondiendo al epidermis y otro al fondo haciendo uso de la operación de *one hot encoding*. El *one hot encoding* consiste en codificar una variable con n estados en n bits de los cuales solo uno toma el valor 1. En este trabajo un canal de las máscaras de etiqueta y de predicción corresponde a la epidermis, entonces los pixeles de epidermis valen 1 y los otros 0 y otro canal corresponde al fondo entonces los pixeles del fondo valen 1 y los otros 0 como se ve en la ilustración 38.

La técnica del *one hot encoding* permite utilizar una codificación de las dos clases de interés (epidermis y fondo) sin que haya una ordenación entre las categorías. Poniendo epidermis a 1 y fondo a 0 el modelo podría considerar una ordenación natural entre las categorías y eso puede dar lugar a malos resultados. El *one hot encoding* permite también pasar dos veces las máscaras de *ground truth* en entrada del modelo para ayudar la convergencia del entrenamiento de la red neuronal convolucional.

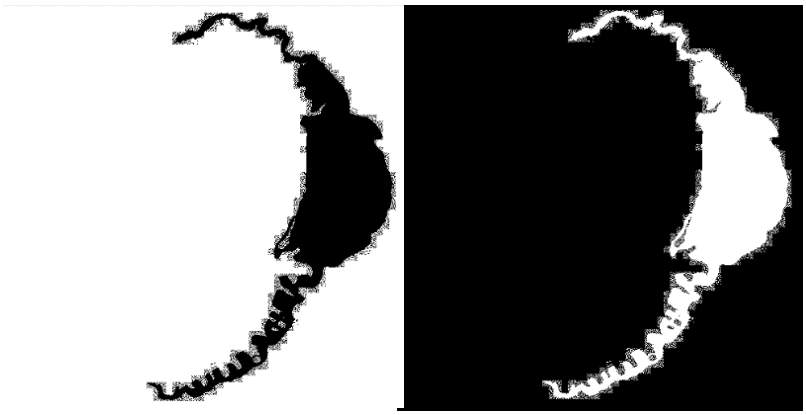


ILUSTRACIÓN 40: EJEMPLO DE LOS DOS CANALES CON *ONE HOT ENCODING* EN ENTRADA DE LA RED (ILUSTRACIÓN PROPIA)

4.7. Técnica de validación cruzada

El entrenamiento se realiza con la técnica de validación cruzada. La validación cruzada es un método de muestreo utilizado para la evaluación del modelo. Es especialmente útil cuando se usa una base de datos bastante limitada en términos de números de datos diferentes. En este trabajo las imágenes de la base de datos de entrenamiento no son muy numerosas y por eso se ha elegido realizar el entrenamiento con esta técnica.

El procedimiento de la validación cruzada consiste en elegir un parámetro k que define el número de grupos en los que se divide la base de datos mezclada aleatoriamente previamente. Se definen k grupos de los cuales (k-1) sirven como datos de entrenamiento y el grupo restante para la validación. Se realizan k iteraciones cambiando cada vez el grupo que sirve para la validación como se ve en la Ilustración 41. Es un método útil para estimar la habilidad de un modelo de aprendizaje automático en datos no vistos durante el entrenamiento. Se obtienen entonces k modelos diferentes para los cuales se conservan las métricas de evaluación del entrenamiento y de la validación. Se escoge el modelo que ofrece mejores resultados en validación.

Es importante guardar el modelo que da mejores resultados en datos que no han servido durante el entrenamiento para paliar a los problemas de *overfitting* que pueden llevar los modelos de redes neuronales profundas con muchas épocas de entrenamiento. El *overfitting* o sobre ajuste en español ocurre cuando el modelo aprende demasiado los datos que sirven al entrenamiento y adapta sus parámetros para predecir a la perfección esos datos y no poder predecir tan bien nuevos datos. Suelen ocurrir los problemas de *overfitting* en bases de datos pequeñas de entrenamiento.

A la hora de implementar este método es importante de reiniciar los pesos de forma aleatoria de la red neuronal entrenada para obtener k modelos independientes. Para hacer eso se utiliza la función Xavier_uniform que rellena un tensor según el método descrito en [51]. Este método utiliza una distribución uniforme que provee un nuevo esquema de inicialización que aporta una convergencia sustancialmente más rápida que una inicialización solo aleatoria.

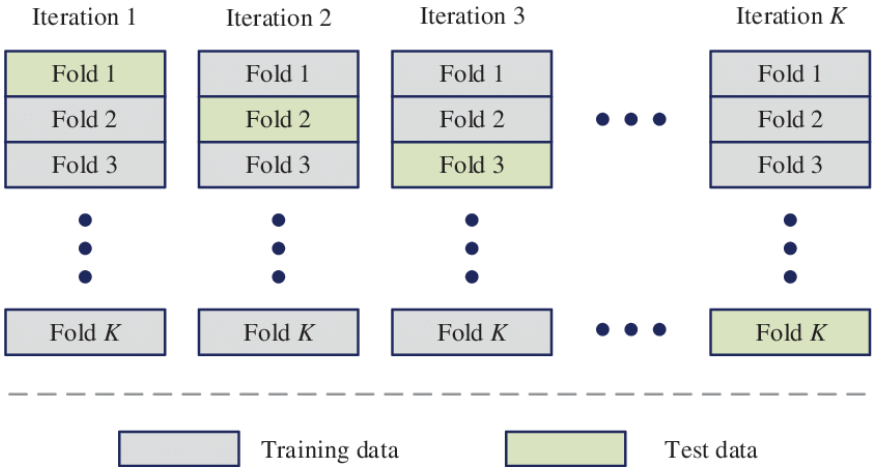


ILUSTRACIÓN 41: ESQUEMA DE LA VALIDACIÓN CRUZADA [51]

En este trabajo se ha elegido un valor de k igual a 5. Con CLARIFYv1, se obtienen 5 grupos de cada uno 11 imágenes de biopsia completa, pero al nivel de los *patches* el número de imágenes en cada grupo es distinto como lo muestra la Tabla 7.

TABLA 7: REPARTICIÓN DE LOS *PATCHES* EN LA PRIMERA VALIDACIÓN CRUZADA

Grupo	**0**	**1**	**2**	**3**	**4**
Numero de *patch*	1 193	920	707	1 251	1 052

5. Resultados

5.1. Resultados de la U-Net para la segmentación gruesa

Estos son los resultados obtenidos del experimento 1 con 150 épocas de entrenamiento en las imágenes de validación:

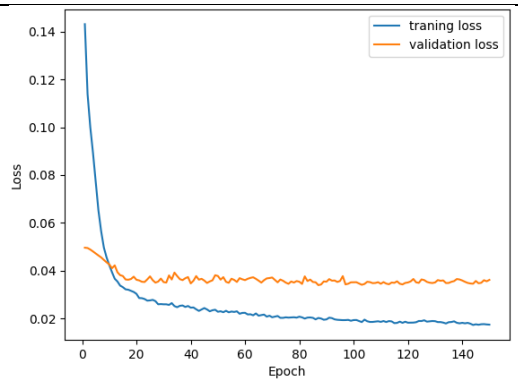


ILUSTRACIÓN 42: GRAFICO DE LA FUNCIÓN DE LA PÉRDIDA EN FUNCIÓN DE LAS ÉPOCAS

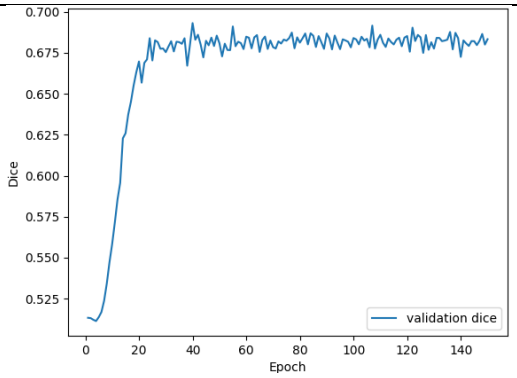


ILUSTRACIÓN 43: MÉTRICA DE *DICE* EN FUNCIÓN DE LAS ÉPOCAS EN EL SET DE VALIDACIÓN

Las curvas de función de pérdidas de la Ilustración 42 y métricas de la Ilustración 43 en función de las épocas del entrenamiento muestran que se realiza de forma correcta el entrenamiento de la U-Net para mejorar los resultados en validación y que el entrenamiento converge. Esta convergencia se hace alrededor de la época 30, como se puede ver en el momento que la función de pérdida deja de bajar en los conjuntos de entrenamiento y de la validación y respectivamente la métrica de *dice* deja de subir de manera significativa. Para incrementar el *dice* de validación se debería aumentar el nivel de detalle de las imágenes procesadas haciendo *patches,* lo que se hace es un segundo tiempo en este trabajo. Se almacena el modelo que obtiene el mejor *dice* en validación.

Con 150 épocas la métrica *dice* es de 0.77 en datos de entrenamiento lo que es alto para una predicción gruesa de la región de la epidermis. Pero el resultado corresponde a la resolución disponible de las imágenes de entrenamiento del modelo.

La Tabla 8 presenta ejemplos de predicción de la epidermis con las imágenes de entrada de la red correspondiente. Son imágenes que provienen de la base de datos CLARIFYv1. Se compara esta predicción con la máscara de *ground truth* correspondiente. Se ven los artefactos previamente comentados dentro de la dermis y que no pertenecen a la epidermis. En cambio se nota que no hay falsos negativos, quiere decir que la red predice bien las zonas de epidermis. El error de predicción que se obtiene es que la máscara es demasiado ancha comparativamente con la epidermis real presente en la imagen. Eso genera ciertos falsos positivos. No hay zonas de epidermis que el modelo no ha sido capaz de detectar en los ejemplos presentados y en la mayoría de los casos de las imágenes de la base de datos CLARIFYv1.

TABLA 8: EJEMPLOS DE PREDICCIÓN OBTENIDA CON IMÁGENES DE TEST

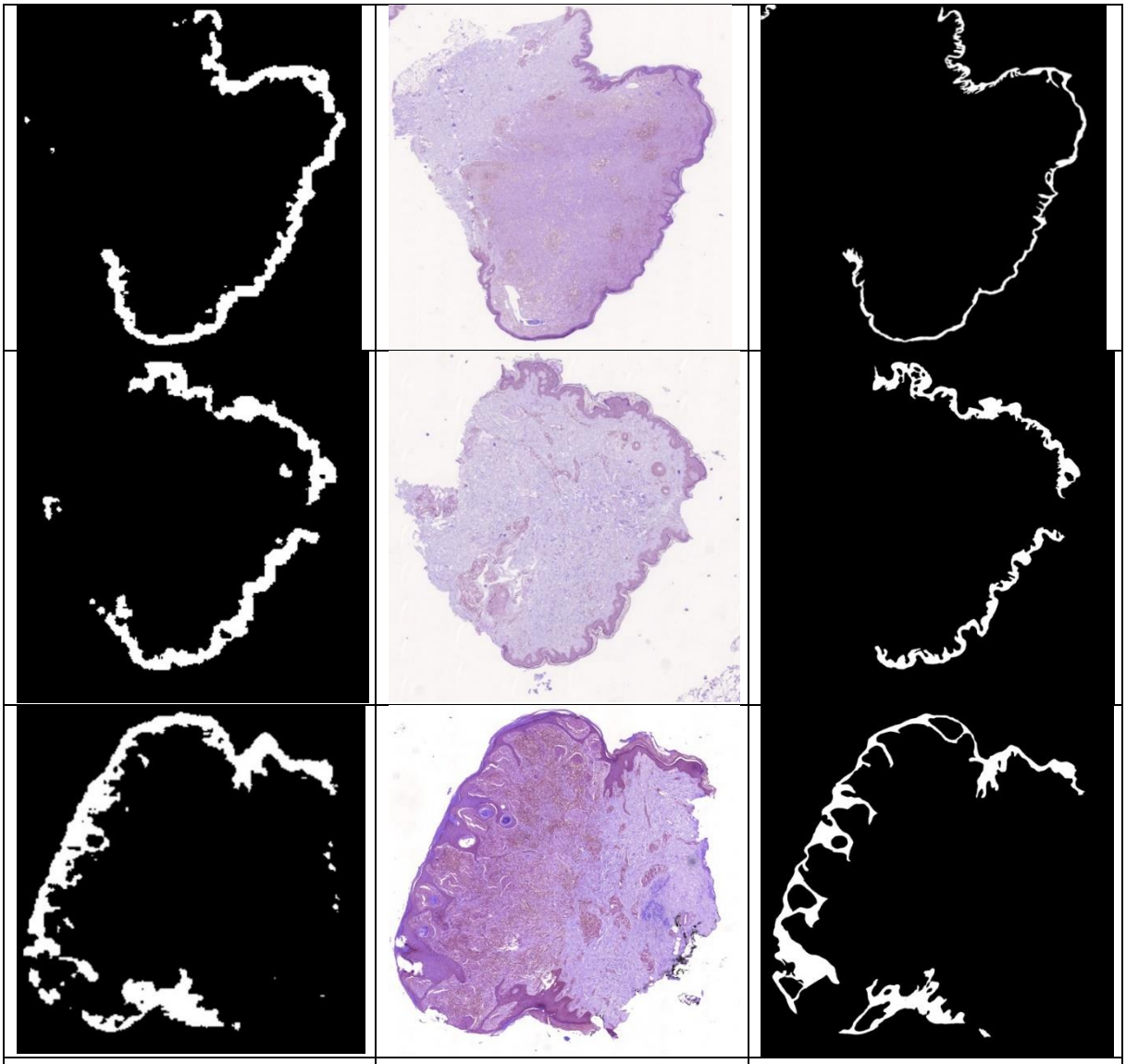


Imagen en salida de la U-Net	Imagen	Mascara de *ground truth* correspondiente

La Tabla 9 a continuación presenta los resultados de la métrica *dice* para los tres conjuntos de datos de la base de datos 1 con la desviación estándar correspondiente. Como otro conjunto de validación externa, se presenta la métrica de *dice* obtenida con la predicción de las imágenes de la base de datos CLARIFYv2. Se obtienen métricas bastante altas sobre todo en entrenamiento, aunque un poco menos para los conjuntos de test y de validación, lo que tiene sentido ya que son imágenes que la red no usa para su entrenamiento. La desviación estándar pequeña asociada permite confirmar que la métrica es alta con poca variabilidad según las imágenes del conjunto de datos en cuestión.

TABLA 9: MÉTRICAS DE *DICE* DEL EXPERIMENTO 1 PARA 150 ÉPOCAS

	U-NET	
	***Dice* score**	**Desviación estándar**
TRAIN	0,7716	0,058
VALIDATION	0,6399	0,010
TEST	0,6387	0,010
CLARIFYV2	0,4876	0,162

La Tabla 10 presenta máscaras de epidermis obtenidas con el modelo en imágenes de la base de datos CLARIFYv2. Esta validación externa del modelo permite ver que el modelo no predice de manera bastante precisa y fiel la epidermis presentando muchos falsos positivos y falsos negativos. Estos resultados pueden indicar que la base de datos número 1 no es suficientemente amplia para obtener un modelo robusto de segmentación de la epidermis en imágenes histopatológicas con lesiones melanocíticas.

TABLA 10: EJEMPLO DE PREDICCIONES OBTENIDAS CON IMÁGENES DE LA BASE DE DATOS CLARIFYV2

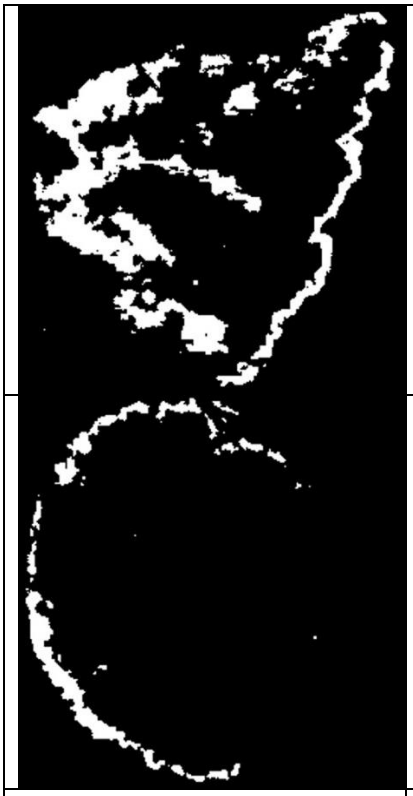
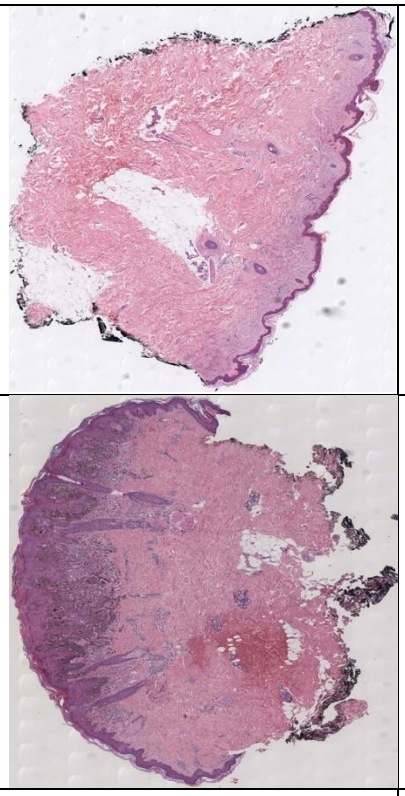
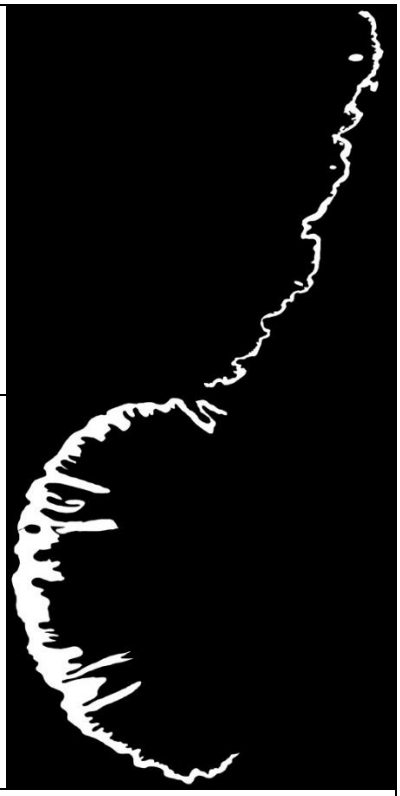
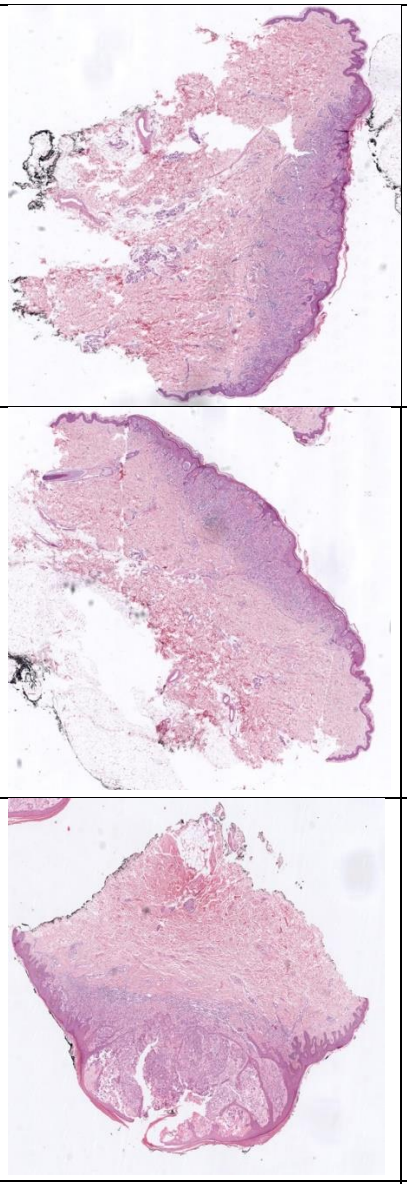
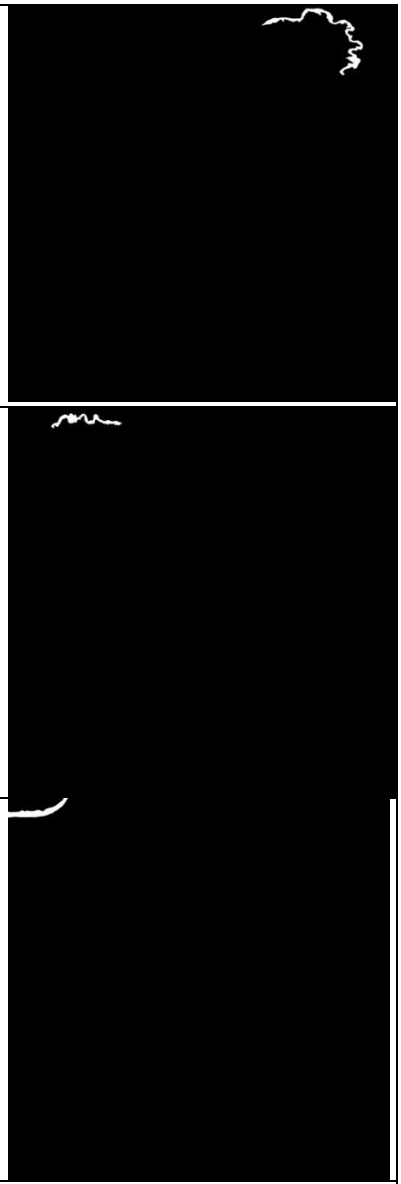
		

Imagen en salida de la U-Net	Imagen H&E	Mascara de *ground truth* correspondiente

Para la inferencia, con las imágenes con máscara mal anotadas se obtiene un *dice* score de 0.52. Es normal tener un valor de la métrica *dice* bajo porque la máscara de *ground truth* no es correcta y también se obtienen muchos **falsos positivos** resultando en una segmentación bastante gruesa de la epidermis como se ve en la Tabla 11. Pero, sí que el modelo implementado lleva a cabo las predicciones correctamente en ciertas imágenes de inferencia que pueden resultar difíciles de predecir. Esta inferencia permite validar la calidad del modelo obtenido.

TABLA 11: RESULTADOS DE INFERENCIA DE LA U-NET EN IMÁGENES CON ANOTACIÓN INCORRECTA

Imagen en salida de la U-Net	Imagen H&E original	Mascara de *ground truth* correspondiente
		

Entonces el experimento 1 permite obtener un primer modelo de segmentación automática de la epidermis en imágenes histopatológicas de piel con lesiones melanocíticas. Pero a un a nivel grueso de la región de la biopsia se puede intentar mejorar los resultados obtenidos eligiendo una arquitectura de red neuronal convolucional en ResU-Net como se presenta en la siguiente sección.

5.2. Resultados de la ResU-Net para la segmentación gruesa

Las curvas de función de pérdida de la Ilustración 44 y métricas de la Ilustración 45 en función de las épocas del entrenamiento muestran que se realiza de forma correcta el entrenamiento de la ResU-Net (experimento 2). Se obtienen una curva de entrenamiento similar a la arquitectura U-Net. Cambia que la convergencia se produce con más épocas, en la época 50, el entrenamiento es por lo tanto un poco más lento respectivamente con el experimento 1. Se considera igual que en el caso precedente el modelo que da los mejores resultados de métricas en el conjunto de validación.

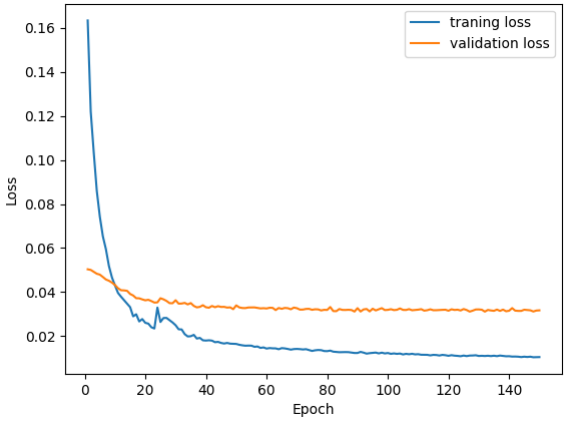


ILUSTRACIÓN 44: GRAFICO DE LA FUNCIÓN DE LA PÉRDIDA EN FUNCIÓN DE LAS ÉPOCAS

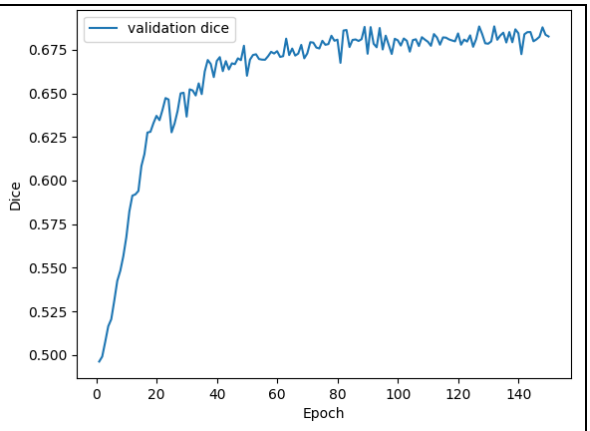


ILUSTRACIÓN 45: MÉTRICA DE *DICE* EN FUNCIÓN DE LAS ÉPOCAS EN EL SET DE VALIDACIÓN

La Tabla 12 a continuación presenta los resultados de la métrica *dice* para los tres conjuntos de datos con la desviación estándar correspondiente y la comparación con los resultados obtenidos en imágenes de la base de datos CLARIFYv2.

TABLA 12: MÉTRICAS DE *DICE* DEL EXPERIMENTO 2 PARA 150 ÉPOCAS

	RESU-NET	
	***Dice* score**	**Desviación estándar**
ENTRENAMIENTO	0,7860	0,050
VALIDACIÓN	0,6491	0,110
TEST	0,6487	0,039
CLARIFYV2	0,5259	0,130

5.3. Comparación de los resultados de la U-Net y de la ResU-Net de los experimentos 1 y 2

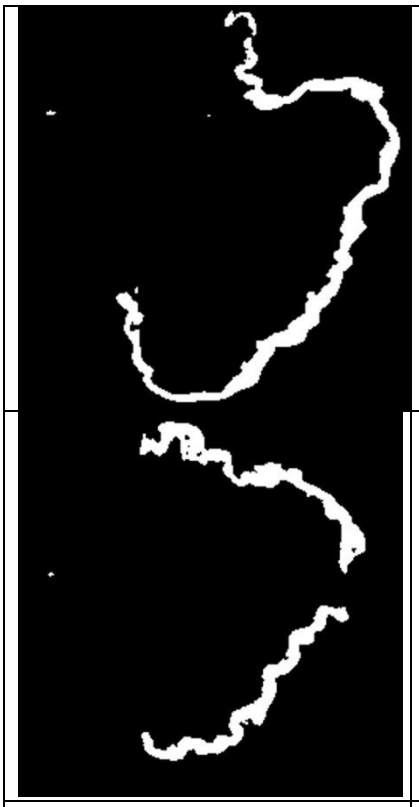
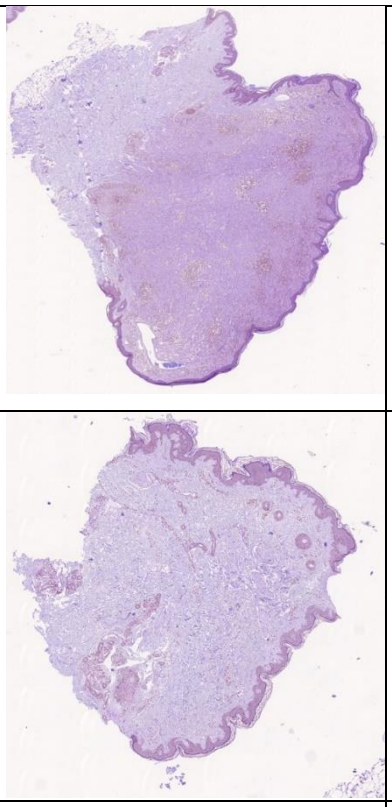
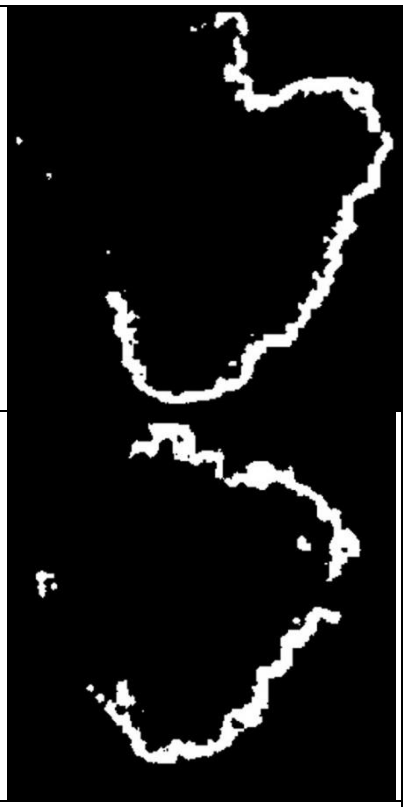
En esta sección se van a comparar los resultados obtenidos entre los dos modelos previamente presentados, primero en término de la métrica de d*ice* como se ve en la Tabla 13. Se ve que el modelo del experimento 2 predice mejor las nuevas imágenes con un *dice* en la base de datos CLARIFYv2 de 0.5259 comparado con un dice de 0.4876 para el experimento 1. Las métricas de *dice* de los conjuntos de imágenes de la base de datos CLARIFYv1 son más elevadas en el caso del experimento 2 comparativamente con el experimento 1 aunque esa diferencia no es tan grande como en el caso de las imágenes de la base de datos CLARIFYv2. La Tabla 13 permite concluir que con los hiperparámetros iguales la arquitectura de red neuronal convolucional de *encoder-decoder* residual mejora los resultados obtenidos para la tarea de segmentación de la epidermis en imágenes histopatológicas con lesiones melanocíticas.

TABLA 13: COMPARACIÓN DE LAS MÉTRICAS DE *DICE* ENTRE LOS EXPERIMENTOS 1 Y 2

	U-NET	RESU-NET
EXPERIMENTO	1	2
ENTRENAMIENTO	0,7716	0,7860
VALIDACIÓN	0,6387	0,6487
TEST	0,6399	0,6491
CLARIFYV2	0,4876	0,5259

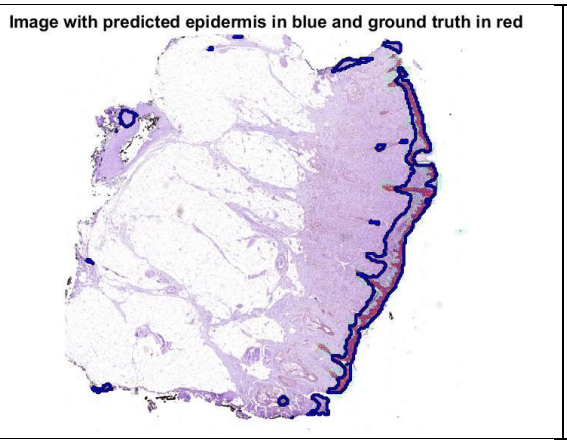
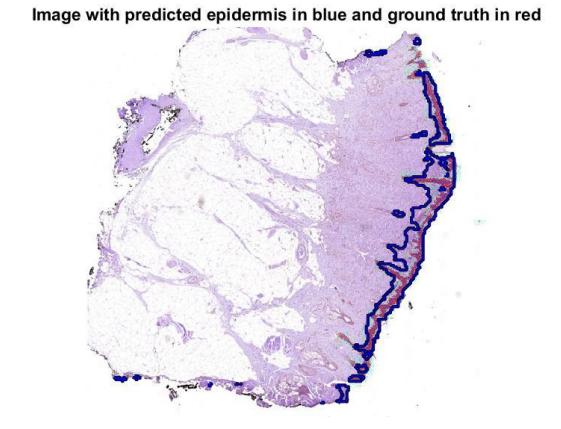
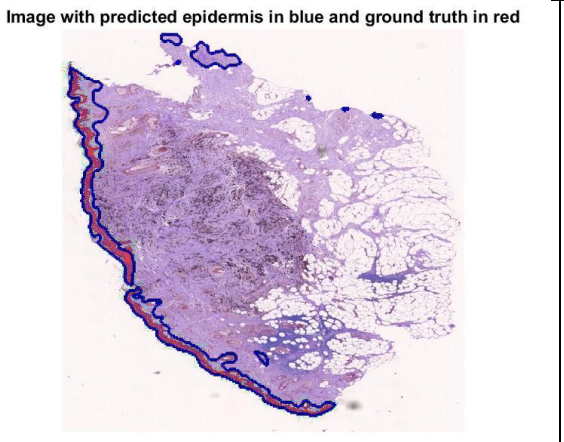
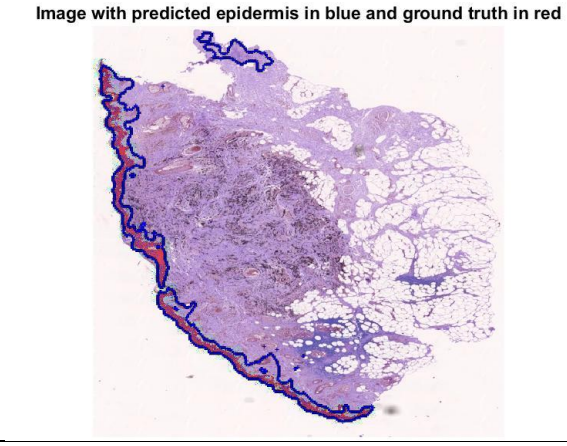
Se comprueba la mejora del experimento 2 con respecto al experimento 1 visualizando las máscaras de predicciones obtenidas. Las máscaras de epidermis obtenidas con el modelo de ResU-Net se aprecian más lisas y precisas que las máscaras obtenidas con la U-Net como se ve en la Tabla 14. Parece que hay menos artefactos que en las máscaras predichas por la arquitectura U-Net y que cuando hay algunos son de menor tamaño. En el caso de la ResU-Net se obtiene también una segmentación más continua de la epidermis y entonces que puede parecer mejor para la mayoría de las imágenes que con la arquitectura U-Net.

TABLA 14: COMPARACIÓN DE LAS MÁSCARAS DE PREDICCIÓN DE LA RESU-NET Y DE LA U-NET EN IMÁGENES DE TEST

		
Máscara obtenida con la Res U-Net	Imagen original	Máscara obtenida con la U-Net

Otra técnica de visualización de los resultados y de comparación de los modelos consiste a realizar el *overlay* de la máscara de segmentación sobre la imagen H&E original. Se pueden visualizar la superposición de la máscara de *ground truth* en rojo y los bordes de la predicción obtenida con el modelo de segmentación en azul como se ve en la Tabla 15. Se obtienen con esta visualización resultados similares entre los dos modelos. Se ven claramente los artefactos dentro de la dermis que habría que quitar y que las predicciones obtenidas tienen un espesor más ancho que la epidermis real.

TABLA 15: COMPARACIÓN DE LOS RESULTADOS ENTRE LA U-NET (EXPERIMENTO 1) Y LA RESU-NET (EXPERIMENTO 2)

	
	
Resultados ResU-Net	Resultados U-Net

5.4. Experimentos a nivel grueso utilizando la técnica de validación cruzada

Para intentar mejorar los resultados previamente obtenidos del modelo a nivel grueso de segmentación de la epidermis (aumentar las métricas de dice y obtener un modelo más robusto), se entrena de nuevo las dos arquitecturas, pero utilizando esta vez la técnica de validación cruzada. Los experimentos que se realizan mediante esta técnica son los de 3 a 8 (ver Tabla 6) con arquitecturas U-Net y ResU-Net alternativamente como se representan en la Tabla 16 para recordatorio.

TABLA 16: HIPERPARÁMETROS DE LOS EXPERIMENTOS 3 A 8 QUE SE REALIZAN CON LA TÉCNICA DE VALIDACIÓN CRUZADA

Nivel de resolución	Experimento	Tipo de arquitectura de red	Base de datos	Validación cruzada	Numero épocas	Learning rate	Batch size	Tamaño imagen
Segmentación gruesa	3	U-Net	1	Si	150	0,0001	16	256
	4	Residual U-Net	1	Si	150	0,0001	16	256
	5	U-Net	2	Si	150	0,0001	16	256
	6	Residual U-Net	2	Si	150	0,0001	16	256
	7	U-Net	1&2	Si	150	0,0001	16	256
	8	Residual U-Net	1&2	Si	150	0,0001	16	256

Usando la técnica de validación cruzada con las WSI redimensionadas de la base de datos CLARIFYv1 se obtienen los resultados de métricas *dice* de la Tabla 17. Se usa la métrica de *dice* en validación para comparar los modelos ya que es una métrica menos sensible a problemáticas de *overfitting* que el *dice* sobre los datos de entrenamiento. Los diferentes grupos de la validación cruzada están detallados en la Tabla 7 de la sección 3.2. Se ve que el grupo 3 del modelo U-Net (experimento 3) presenta los mejores resultados y el grupo 2 del modelo ResU-Net (experimento 4) respectivamente. Son entonces los dos modelos que se van a guardar como mejores para realizar las pruebas de inferencia siguientes. En ambos casos la métrica de *dice* de validación obtenida es mayor que la obtenida sin la implementación de la técnica de validación cruzada en el proceso de entrenamiento.

TABLA 17: RESULTADOS DE *DICE* DE EXPERIMENTO CON Y SIN VALIDACIÓN CRUZADA CON LOS MODELOS DE U-NET Y RESU-NET CON IMÁGENES DE LA BASE DE DATOS CLARIFYV1

	Experimento			
	3	**4**	**1**	**2**
CLARIFYv1	**validación**	**validación**	**validación**	**validación**
grupo	*dice*	*dice*	*dice*	*dice*
0	0,7004	0,6961	0,6399	0,6491
1	0,6979	**0,7218**		
2	0,6953	0,6999		
3	**0,7020**	0,6963		
4	0,6802	0,6872		

A continuación, se comparan las métricas obtenidas por los diferentes modelos de los experimentos 1 a 4 en el conjunto de test de la base de datos CLARIFYv1 y sobre los datos de validación externa de la base de datos CLARIFYv2. En ambos casos son datos que no han servido durante el entrenamiento y tampoco en validación del entrenamiento. Las métricas obtenidas sobre los datos de inferencia de la base de datos CLARIFYv2 son importantes a la hora de evaluar los modelos ya que proveen de otro set de datos y permiten una validación más robusta.

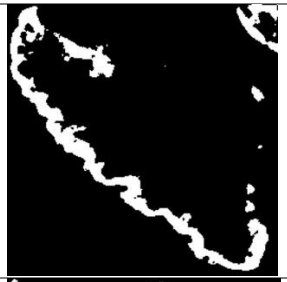
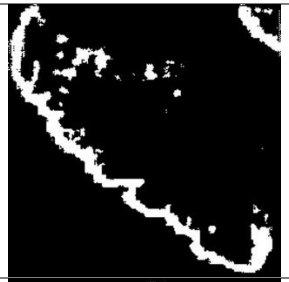
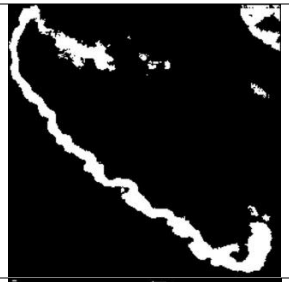
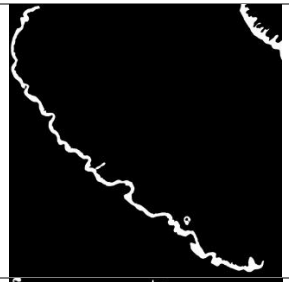
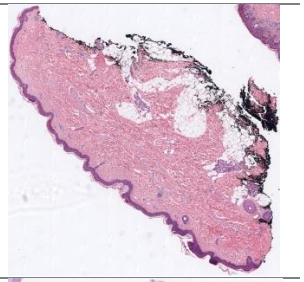
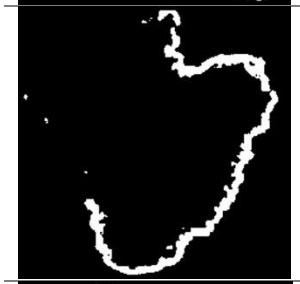
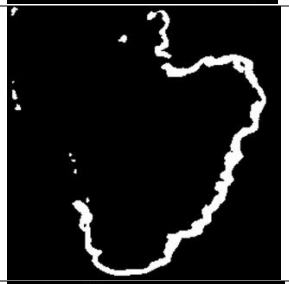
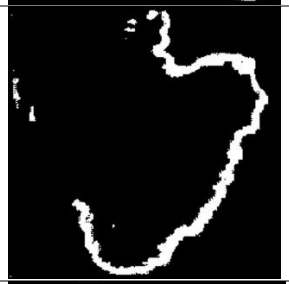
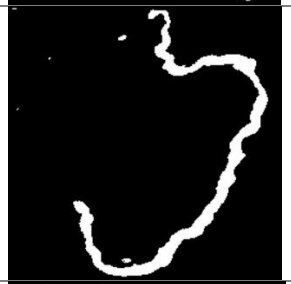
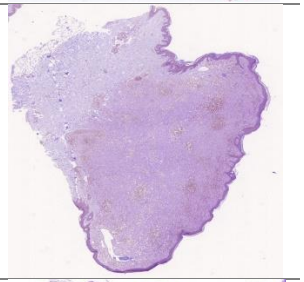
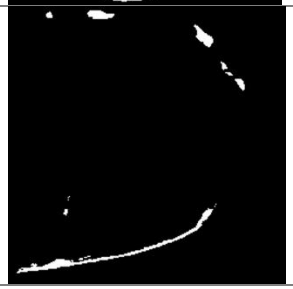
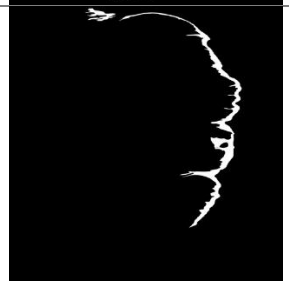
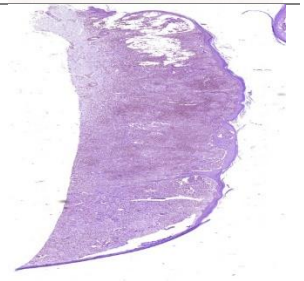
La Tabla 18 presenta las métricas de *dice* que se obtienen. Los diferentes modelos sacan todos una métrica de *dice* parecida de 0.62 en promedio sobre las imágenes de test de la base de datos CLARIFYv1 y de 0.51 sobre las imágenes de la base de datos CLARIFYv2. En estos casos la mejora obtenida con la técnica de validación cruzada no es tan evidente como anteriormente. Las métricas de *dice* en imágenes de test de la base de datos CLARIFYv1 son más bajas usando los modelos obtenidos con validación cruzada. En cambio sí que en ambos casos la métrica de *dice* en imágenes de la base de datos CLARIFYv2 es más alta. Entonces en el caso de las imágenes completas, el método de validación cruzada mejora las métricas y predicción durante el entrenamiento pero el resultado en conjunto de test es similar a un modelo entrenado sin validación cruzada.

TABLA 18: RESULTADOS DE *DICE* EN DATOS DE TEST PARA LOS MODELOS U-NET Y RESU-NET CON Y SIN VALIDACIÓN CRUZADA

	Validación cruzada		**Sin validación cruzada**	
	U-Net	**ResU-Net**	**U-Net**	**ResU-Net**
Experimento	3	4	1	2
test CLARIFYv1	0,6108	0,6089	0,6387	0,6487
CLARIFYv2	0,5101	0,5295	0,4876	0,5259

La Tabla 19 presenta ejemplos de imágenes de mascara obtenidas con los diferentes modelos de los experimentos 1 a 4. La primera fila corresponde a una imagen de la base de datos CLARIFYv2 y las otras filas a imágenes de la base de datos CLARIFYv1 sobre la cual los modelos se han entrenados. Pero en todos casos son imágenes de test para evaluar visualmente la robustez del modelo y poder validarlo. Las dos primeras filas presentan resultados similares entre los diferentes experimentos, se localiza correctamente la zona de la epidermis aunque su segmentación sigue siendo gruesa con una máscara espesa y poco precisa. Conforme la idea que este modelo es útil como primera etapa para seleccionar los *patches* de interés dentro de la imagen muy grande que es la biopsia completa. Según la imagen los modelos predicen una máscara con menor o mayor ruido pero en la mayoría de los casos se ve artefactos de ruido que empeoran el resultado deseado. La tercera fila de la tabla presenta una imagen difícil de predecir por todos los modelos, en este caso el experimento 3 es el que da mejor resultado ya que se seleccionan las zonas blancas para sacar los *patches* y obtener la segmentación de precisión. Los otros experimentos que presentan muchos falsos negativos para esta imagen en particular implicarían seleccionar zonas que en realidad no pertenecen al *ground truth* y no seleccionar las zonas correctas, con lo cual el modelo de precisión no sería capaz de obtener una segmentación fina y correcta de la epidermis de estas imágenes con lesiones melanocíticas.

TABLA 19: EJEMPLO DE MÁSCARAS DE PREDICCIÓN DE LOS EXPERIMENTOS 1 A 4 DE DATOS DE TEST DE LAS DOS BASES DE DATOS

Experimento				Ground truth	Imagen original
1	2	3	4		
					
					
					

Para seguir intentando mejorar los resultados del modelo se elige añadir las imágenes nuevamente anotadas de la base de datos CLARIFYv2 a los datos de entrenamiento. Se tratan de los experimentos 5 a 8 de la Tabla 16 (ver también Tabla 6). Las métricas obtenidas en los diferentes conjuntos de datos se presentan en la Tabla 20.

Los modelos 5 y 6 presentan resultados muy buenos en los conjuntos de validación y test de la base de datos CLARIFYv2. En cambio, los resultados de la métrica de *dice* sobre la base de datos CLARIFYv1 empeoran. Este fenómeno se puede explicar con el hecho que los modelos de los experimentos 5 y 6 han sido entrenados únicamente con imágenes que pertenecen a la base de datos CLARIFYv2.

Los dos últimos experimentos tienen en consideración las dos bases de datos disponibles. Sus entrenamientos se realizan cargando los pesos de la red neuronal convolucional de los experimentos 3 y 4 y reentrenando el modelo sobre imágenes de la base de datos CLARIFYv2. Este procedimiento se llama *fine-tuning* y se detalla a continuación.

Para acelerar el entrenamiento se considera cargar en memoria los pesos de una red U-Net que segmenta la epidermis de los experimentos 3 y 4 y entrenarla de nuevo pero solo haciendo más precisos los pesos con las imágenes de la base de datos CLARIFYv2. El *fine-tuning* es una técnica parecida al *transfer learning* que consiste en cargar los pesos de una red preentrenada sobre un conjunto bastante amplio de imágenes para entrenar de nuevo la red y que se afine para realizar la segmentación de interés.

Este procedimiento permite que los experimentos 7 y 8 presenten métricas de *dice* altas en los diferentes conjuntos de datos de las dos bases de datos. Comparando las diferentes medias de la Tabla 20 se ve que son más altas en todos los casos presentes. Esta tabla permite también comparar los modelos con arquitecturas U-Net (5 y 7) y ResU-Net (6 y 8). En los dos casos (experimento 5 comparado con el experimento 6 y experimento 7 comparado con experimento 8), los modelos con arquitectura de red neuronal convolucional *encoder-decoder* con capas residuales (ResU-Net) presentan mejores resultados que los modelos con arquitectura tradicional de U-Net.

El mejor modelo de segmentación gruesa es el modelo 8 que alcanza un *dice* de 0.57 y 0.72 en datos de validación de las bases de datos 1 y 2 y un dice de 0.54 y 0.73 en datos de test de las bases de datos 1 y 2.

TABLA 20: RESULTADOS DE LA MÉTRICA DE DICE EN LOS DIFERENTES CONJUNTOS DE DATOS DE LAS DOS BASES DE DATOS PARA LOS ENTRENAMIENTOS DE 5 A 8

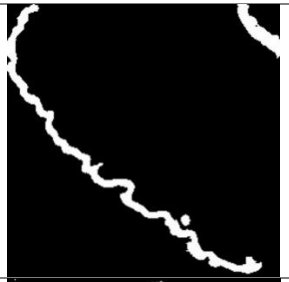
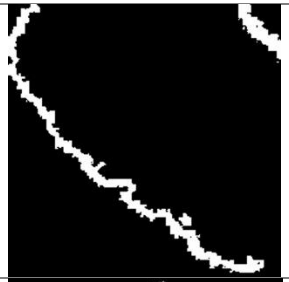
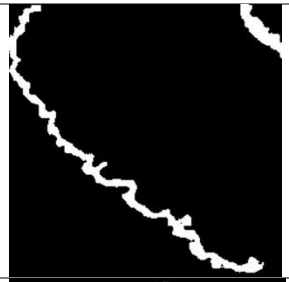
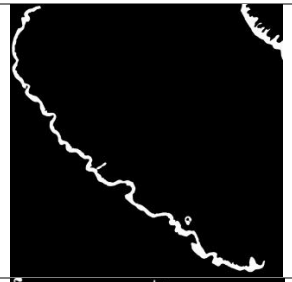
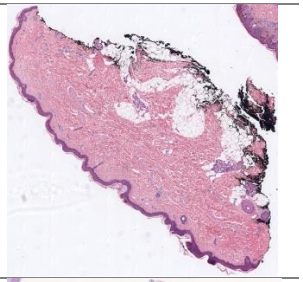
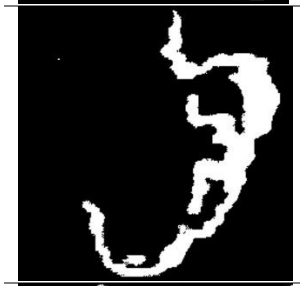
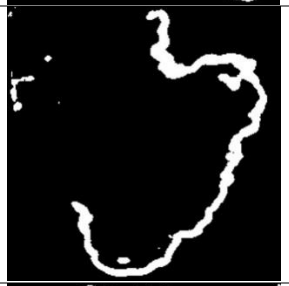
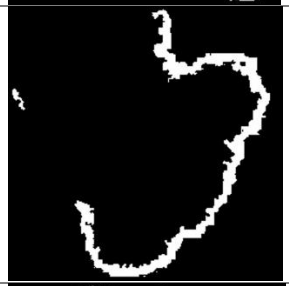
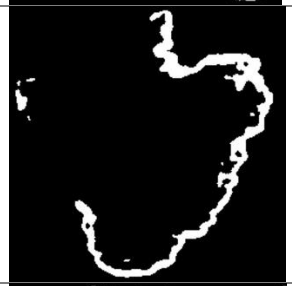
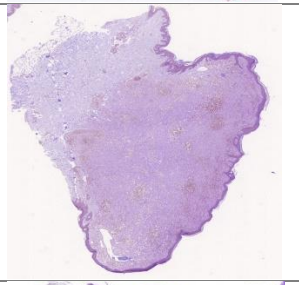
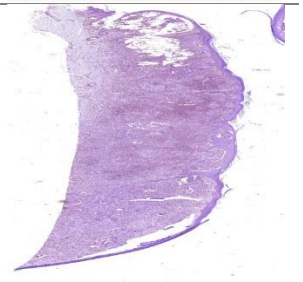
Experimento	Dice en promedio										
	Entrenamiento			Validación			Test			CLARIFYv1	CLARIFYv2
Base de datos	1	2	media	1	2	media	1	2	media		
5	0,4465	0,5205	0,4835	0,4941	0,7141	0,6041	0,4808	0,6881	0,5845	0,4738	0,6409
6	0,4835	0,5556	0,5195	0,5483	0,7204	0,6343	0,5156	0,7229	0,6193	0,5158	0,6663
7	0,5371	0,5722	0,5546	0,5879	0,7024	0,6451	0,5712	0,6960	0,6336	0,5654	0,6569
8	0,5837	0,5949	0,5893	0,5743	0,7272	0,6507	0,5454	0,7339	0,6397	**0,5678**	**0,6853**

La Tabla 21 presenta ejemplos de imágenes de mascara obtenidas con los diferentes modelos de los experimentos 5 a 8. Se visualizan las mismas imágenes que las de la Tabla 19.

La primera fila presenta resultados muy similares entre los diferentes experimentos. En la siguiente fila de la tabla, se ve la mejora de la predicción de una imagen de la base de datos CLARIFYv1 entre los diferentes experimentos. La máscara se hace más fina, precisa y correcta con menos

artefactos de ruido entre los diferentes experimentos. La tercera imagen presenta una máscara de predicción bastante diferentes del *ground truth* en todos los casos. El experimento 5 es el que presenta los peores resultados. Sin embargo, la ventaja de los experimentos 7 y 8 con respecto a los 3 y 4 previamente presentados es que hay menos falsos negativos. Entonces se van a seleccionar un número más grande de *patches* pero estos *patches* contienen los que realmente pertenecen a la clase de epidermis. Son modelos de segmentación gruesa que son más prometedores como primera etapa de selección de los *patches* de interés para luego obtener la segmentación precisa de la epidermis de imágenes histopatológicas con lesiones melanocíticas.

TABLA 21: EJEMPLO DE MÁSCARAS DE PREDICCIÓN DE LOS EXPERIMENTOS 5 A 8 DE DATOS DE TEST DE LAS DOS BASES DE DATOS

Experimento				Ground truth	Imagen original
5	6	7	8		
					
					
					

5.5. *Patches* de la imagen y resultados de la segmentación precisa

Con las máscaras de predicción de las epidermis obtenidas se pueden realizar *patches* de las regiones de las imágenes originales para aumentar el nivel de detalle de la predicción realizada. Se entrena un nuevo modelo de U-Net con la misma arquitectura que presentada previamente pero con imágenes de entrada *patches* de tamaño 512*512 de la imagen original y un *batch size* de 4 durante 50 épocas correspondiendo al experimento 9 también detallado en la Tabla 6. Como presentado en la ilustración 31, se dispone de un numero mucho más grande de imágenes a nivel de *patches* que a nivel de imagen completa. Hay lo suficiente de datos para que no haga falta utilizar procedimientos como la validación detallado en la sección 4.7.

Las curvas de función de pérdida de la Ilustración 46 y métricas de la Ilustración 47 en función de las épocas del entrenamiento muestran que se realiza de forma correcta el entrenamiento de la U-Net. En este caso la convergencia se alcanza con pocas épocas, solo con 10 ya se ha obtenido la convergencia deseada. Las métricas *dice* en el conjunto de validación son también más altas que en los experimentos precedentes de más baja resolución.

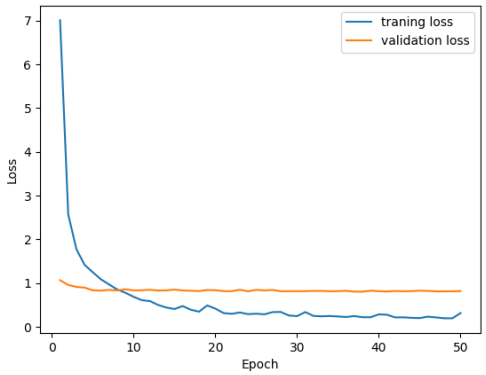


ILUSTRACIÓN 46: GRAFICO DE LA FUNCIÓN DE PÉRDIDAS EN FUNCIÓN DE LAS ÉPOCAS

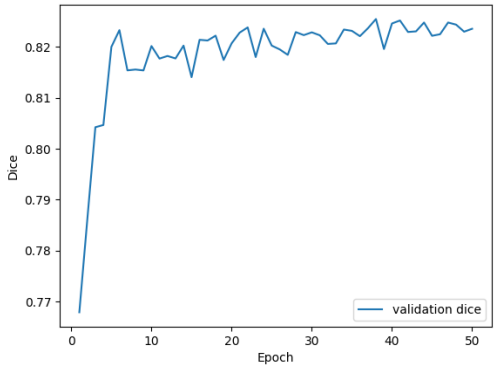
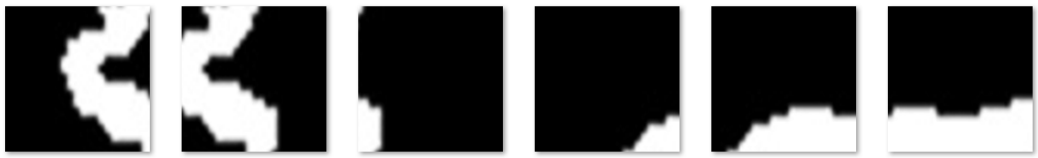
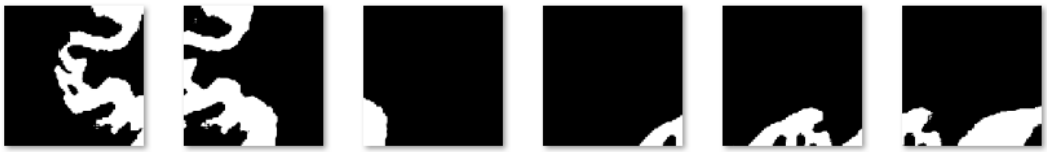
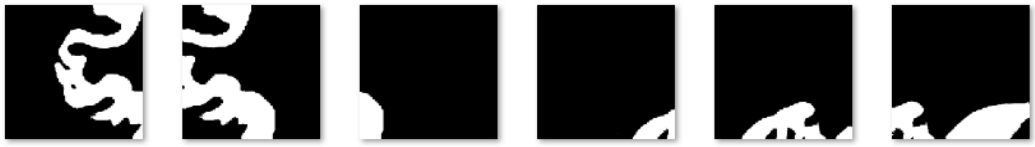


ILUSTRACIÓN 47: GRAFICO DE LA MÉTRICA *DICE* EN FUNCIÓN DE LAS ÉPOCAS

La Tabla 22 presenta ejemplo de *patches* con la máscara de *ground truth* correspondiente y las predicciones de las dos arquitecturas a nivel de *patch*. Se ve claramente que la resolución de los *patches* de *ground truth* no es tan alta como la de las predicciones realizadas por los algoritmos lo que puede impactar las métricas de resultados obtenidos. Las máscaras de *ground truth* de la base de datos CLARIFYv1 son de menor resolución y eso penaliza a la hora de obtener las métricas. Las máscaras de *ground truth* de la base de datos CLARIFYv2 no presentan este problema.

TABLA 22: EJEMPLO DE *PATCHES* DE LAS IMÁGENES DE TEST

Imágenes H&E	
Mascara de *ground truth*	
Mascara de predicción de la U-Net	
Mascara de predicción de la ResU-Net	

En la Tabla 23 se presentan los resultados de la comparación de las arquitecturas de U-Net y ResU-Net de los experimentos 9 y 10 (ver Tabla 6). Antes de realizar estos experimentos se ha probado con otros valores del *batch size* y del *learning rate*. Estas pruebas sirven de justificación empírica del ajuste de los hiperparámetros de entrenamiento. Los valores escritos en la Tabla 6 son los que permiten una convergencia de menor duración y con buenas métricas de resultados.

En termino de métricas, la arquitectura de red convolucional *encoder-decoder* U-Net alcanza resultados de *dice* más altos en el conjunto de entrenamiento que la ResU-Net en el caso de una imagen de entrada de tamaño 512. Sin embargo, en los conjuntos validación y test la ResU-Net alcanza mejores resultados. Eso indica que la arquitectura de ResU-Net ofrece un modelo más robusto y con una mejor validación.

TABLA 23: COMPARACIÓN DE LOS RESULTADOS DE LOS EXPERIMENTOS 9 Y 10

EXPERIMENTO	**9**	**10**
DICE	**U-Net**	**ResU-Net**
ENTRENAMIENTO	0,8284	0,7493
TEST	0,7399	0,7488
VALIDACIÓN	0,8554	0,8590

5.6. Resultados de los diferentes experimentos de segmentación precisa

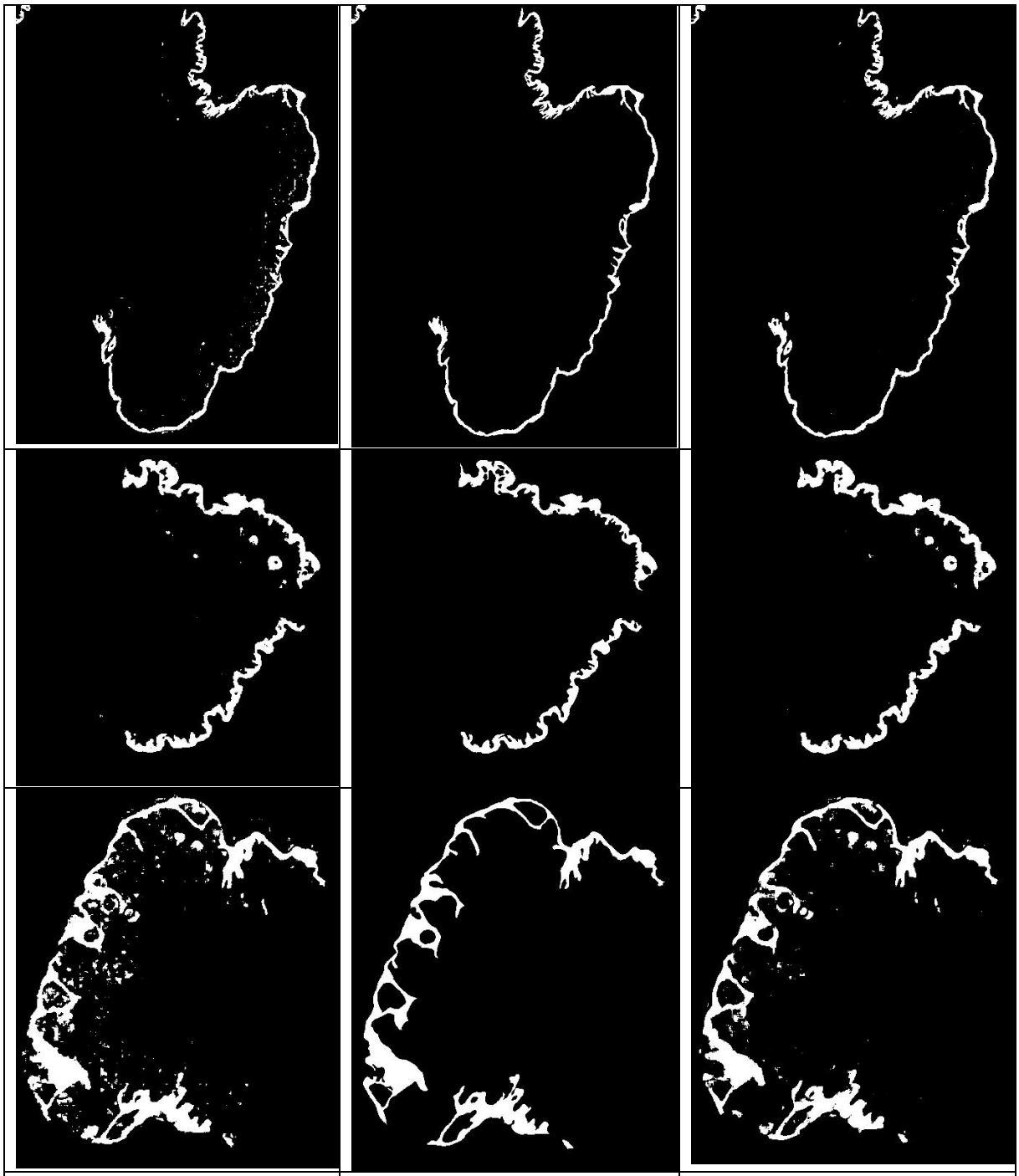
5.6.1. Reconstrucción de las imágenes completas

Para visualizar y comparar los resultados de los experimentos 9 y 10, se realiza la reconstrucción de la imagen completa, juntando todos los *patches* de predicción.

La Tabla 24 presenta imágenes reconstruidas con la U-Net y la ResU-Net de los experimentos 9 y 10 respectivamente. Se comparan con el *ground truth* y se presentan en esta tabla imágenes del conjunto de test. Se aprecia que las imágenes reconstruidas mediante los *patches* predichos de la U-

Net presentan más artefactos y falsos positivos que no pertenecen a la epidermis que los resultados obtenidos mediante la ResU-Net. Además, se observa que en ambos casos las predicciones obtenidas resultan mucho más precisas que las obtenidas en la primera red con las imágenes totales (ver Tabla 19 y Tabla 21). El resultado obtenido es una máscara más fina que segmenta mejor la estructura de la epidermis en imágenes histopatológicas de piel. La última imagen de la tabla siguiente sigue siendo un reto para segmentarla sin falsos positivos y artefactos de ruido.

TABLA 24: COMPARACIÓN DE LAS MÁSCARAS RECONSTRUIDAS OBTENIDAS CON LA U-NET Y LA RESU-NET COMPARATIVAMENTE CON EL GROUND TRUTH



Resultados de la U-Net	*Ground truth*	Resultados de la ResU-Net

Hay unos artefactos en las imágenes de predicción que corresponden a ruido introducido por la red neuronal. Para remediar a este problema se lleva a cabo una etapa de post procesado de las imágenes de salida de la red . Se puede usar la función *remove_small_objects* del paquete *morphology* de la librería *skimage* en Python. Esta función elimina los objetos más pequeños que el tamaño especificado y se le precisa también la conectividad usada que define la vecindad de un píxel. Es una operación morfológica parecida a una apertura. La apertura (*opening*) es una erosión seguida de una dilación en tratamiento de las imágenes como se representa en la Ilustración 48. La apertura tiene como efecto de eliminar las partículas pequeñas (que es el efecto deseado en esta etapa de post procesado) pero puede también separar los objetos más grandes en los puntos donde se estrechan lo que puede resultar malo con respecto al resultado esperado. Por lo tanto, se trata de usar cuidosamente las etapas de post procesado evaluando mediante las métricas la mejor obtenida.

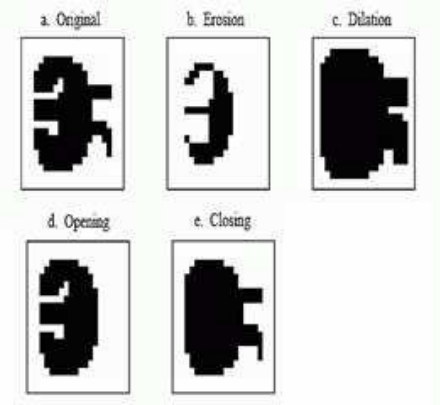


ILUSTRACIÓN 48: OPERACIONES MORFOLÓGICAS SOBRE IMÁGENES [52]

Empíricamente se elige un tamaño máximo de 5000 pixeles para los objetos que se quitan en las etapas de post procesado para eliminar los artefactos de ruido de forma eficaz sin quitar la información relevante como se puede ver en la Ilustración 49. El tamaño mínimo del objecto conservado puede variar de una imagen a otra. Por lo tanto, introduce menos sesgo de no utilizar etapas de post procesado pero esta etapa en concreto permite mejorar bastante los resultados como lo vamos a demostrar a continuación.

ILUSTRACIÓN 49: RESULTADO DEL POST PROCESADO DE QUITAR PEQUEÑOS OBJETOS (A LA IZQUIERDA LA IMAGEN SIN PROCESAR Y A LA DERECHA LA IMAGEN FINAL)

Se computa la métrica dice entre las imágenes reconstruidas y el *ground truth* para comparar con los resultados de los mejores experimentos de segmentación gruesa (experimento 7 y 8 de la Tabla 6) que tienen como entrada las imágenes completas. Los resultados obtenidos se presentan en la Tabla 25 donde se puede apreciar la notable mejora de los resultados con las máscaras reconstruidas a partir de los *patches* comparativamente con las máscaras de salida de los primeros modelos de baja resolución considerando las dos arquitecturas distintas de red neuronal convolucional *encoder-decoder*.

En el conjunto de test, para los modelos a nivel de patch, la U-Net con capas residuales (ResU-Net del experimento 10) tiene una métrica más alta que la U-Net clásica (experimento 9). En los otros conjuntos de datos las métricas son similares entre las dos arquitecturas. Las métricas confirman que la segmentación es mejor con la ResU-Net así como se ha podido constatar en las imágenes previas. La ResU-Net parece proporcionar una buena arquitectura de red neuronal convolucional para la segmentación de la epidermis en imágenes de biopsias de piel humanas con lesiones melanocíticas y con tinción hematoxilina eosina.

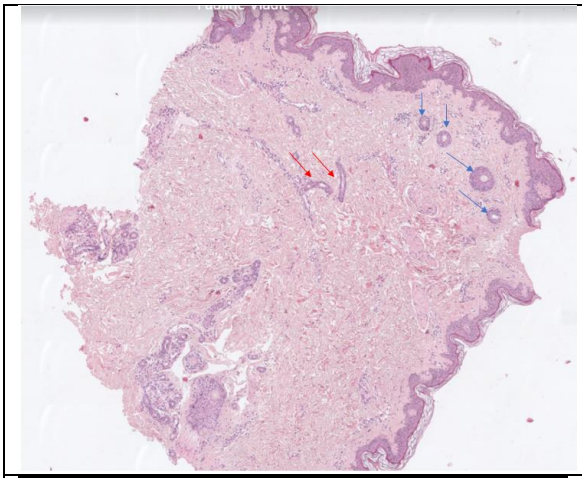
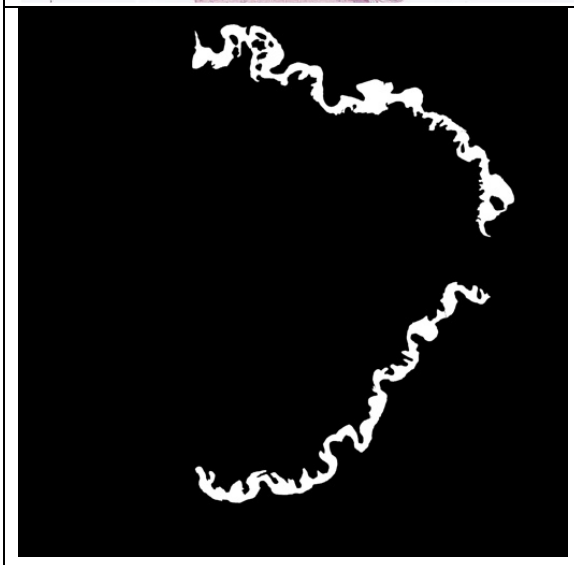
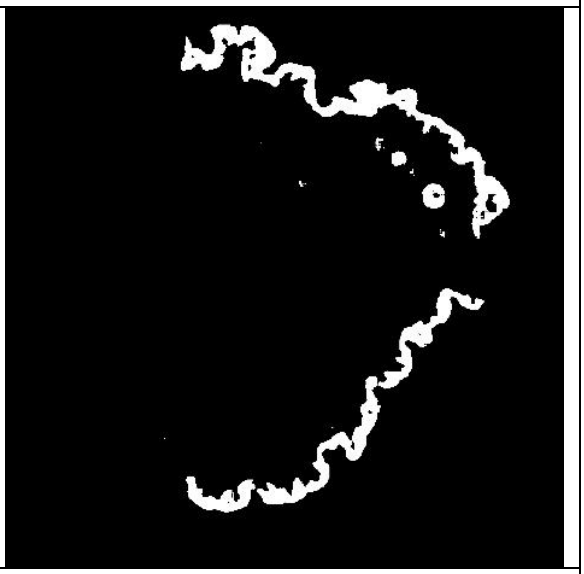
TABLA 25: RESULTADOS DE COMPARACIÓN ENTRE LOS DIFERENTES EXPERIMENTOS SOBRE IMÁGENES DE LA BASE DE DATOS CLARIFYV1

EXPERIMENTO	**7**		**9**		**8**		**10**	
ARQUITECTURA DE LA RED NEURONAL CONVOLUCIONAL	**U-Net gruesa**		**U-Net precisa**		**ResU-Net gruesa**		**ResU-Net precisa**	
MÉTRICA	***dice***	**std**	***dice***	**std**	***dice***	**std**	***dice***	**std**
ENTRENAMIENTO	0,5371	0,16	0,9607	0,01	0,5837	0,16	0,8652	0,09
VALIDACIÓN	0,5879	0,13	0,8539	0,07	0,5743	0,18	0,8533	0,09
TEST	0,5712	0,13	0,7008	0,07	0,5454	0,15	0,7430	0,06

5.6.2. Estructuras detectadas por el modelo

Los resultados de las máscaras reconstruidas de la epidermis presentan estructuras que no están presentes en las máscaras de *ground truth*. Pero observando las estructuras en cuestión no se parecen a artefactos típicos de ruido ya que son estructuras más grandes y con una forma similar a una estructura biológica. Se ha contactado con un patólogo para obtener más información sobre esta detección del modelo. La hipótesis del estudiante y de sus tutores era que las estructuras detectadas dentro de la dermis se correspondían a folículos pilosos que tienen como composición un tejido muy similar al de la epidermis y pueden entonces considerarse parte de ella. Una visualización del problema descrito se ve en la Tabla 26 en la máscara de predicción de la ResU-Net precisa (experimento 10) con respecto a la imagen de ground truth y la imagen H&E de la biopsia original.

TABLA 26: MASCARAS DE DUDAS EN CUENTO A LOS FOLÍCULOS PILOSOS

	Imagen H&E original con las flechas rojas que corresponden a glándulas sudoríparas y las flechas azules al infundíbulo del folículo piloso
	
Mascara de *ground truth*	Mascara de predicción de la ResU-Net a nivel de *patch* (experimento 10)

El patólogo ha confirmado que las estructuras detectadas por la red corresponden al infundíbulo del folículo piloso.

Otras estructuras presentes en las imágenes pero más en profundidad en la dermis corresponden a las glándulas sudoríparas que no pertenecen a la epidermis y que secretan el sudor. Por lo tanto, la red es capaz de realizar la segmentación de la parte de los folículos pilosos que pertenece a la epidermis aunque no estaba presente en las máscaras de anotación.

Así la red obtenida permite detectar en ciertos casos más "evidentes" como el de la Tabla 26 estructuras de epidermis dentro de la dermis pero tampoco es un objetivo de la red implementada. Además al nivel patológico de las lesiones melanocíticas y en particular del melanoma spitzoide, el diagnostico se realiza con información de la epidermis y de su unión con la dermis pero no al nivel del infundíbulo del folículo piloso. Sin embargo, esta estructura podría ser aplicada en otras patologías y el modelo de interés en estos casos específicos.

Este razonamiento permite deducir que los círculos detectados en la Tabla 26 no son errores de segmentación ni debidos al ruido y que las métricas respectivamente al ground truth son en promedio correctas pero a considerar con cuidado a nivel de ciertas imágenes en concreto.

5.6.3. Validación con la base de datos CLARIFYv2

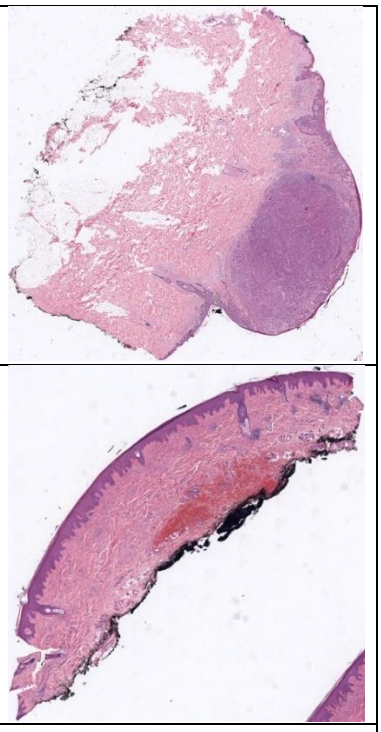
Los modelos obtenidos en los experimentos 9 y 10 solo han sido entrenados sobre imágenes de la base de datos CLARIFYv1. La base de datos CLARIFYv2 puede entonces ser útil para validar los modelos obtenidos. La Tabla 27 presenta los resultados obtenidos con estos experimentos realizando la predicción de las imágenes de la base de datos CLARIFYv2 y se ve que las métricas son bastante altas, del mismo orden de magnitud que las métricas obtenidas en el conjunto de test de la base de datos CLARIFYv1 como se ve en la Tabla 23.

TABLA 27: RESULTADOS DE LA MÉTRICA DE DICE DE LOS EXPERIMENTOS 9 Y 10 SOBRE DATOS DE VALIDACIÓN EXTERNA DE LA BASE DE DATOS CLARIFYV2

EXPERIMENTO	9		10	
ARQUITECTURA	U-Net		ResU-Net	
METRICA	dice	std	dice	std
CLARIFYV2 NIVEL DE PATCHES	0,7184	0,11	0,7289	0,12
CLARIFYV2 IMAGENES RECONSTRUIDAS	0,7029	0,09	0,6768	0,14
CON POST PROCESADO	0,7228	0,09	0,6956	0,14

La Tabla 28 presenta ejemplos de imágenes de las predicciones obtenidas mediante la reconstrucción basadas en *patches* de la base de datos CLARIFYv2. Se ve que las máscaras de epidermis obtenidas segmentan de manera correcta y precisa la epidermis pero con ciertos artefactos de ruido que generan falsos positivos. En cambio, no hay muchos falsos negativos en estos ejemplos. Se puede notar también como un desplazamiento entre la máscara de predicción reconstruida y la máscara de *ground truth*. Este desplazamiento se debe a la división en *patches* de la imagen, como son *patches* de 512*512, el tamaño de la imagen no es siempre un múltiple de 512 y eso genera un desplazamiento durante la reconstrucción.

TABLA 28: EJEMPLO DE MASCARA DE PREDICCIÓN RECONSTRUIDA DEL EXPERIMENTO 9 EN IMÁGENES DE LA BASE DE DATOS CLARIFYV2 CON RESPECTO AL *GROUND TRUTH* Y LA IMAGEN ORIGINAL

		
Mascara de prediccion de la epidermis del experimento 9	Ground truth	Imagen original H&E

5.6.4. Otros experimentos a nivel de *patch*

Para intentar mejorar las métricas obtenidas con los modelos de segmentación de precisión se realizan los experimentos 11 a 14 como se ven en la Tabla 29.

TABLA 29: HIPERPARÁMETROS DE LOS EXPERIMENTOS 9 A 14 QUE SE REALIZAN CON A NIVEL DE PATCH

Nivel de resolución	Experimento	Tipo de arquitectura de red	Base de datos	Validación cruzada	Numero épocas	Learning rate	Batch size	Tamaño imagen
Segmentación precisa	9	U-Net	1	No	50	0,0001	4	512
	10	Residual U-Net	1	No	50	0,0001	4	512
	11	U-Net	2	No	50	0,0001	4	512
	12	Residual U-Net	2	No	50	0,0001	4	512
	13	U-Net	1&2	No	50	0,0001	4	512
	14	Residual U-Net	1&2	No	50	0,0001	4	512

Los resultados obtenidos a nivel de patch se muestran en la Tabla 30. En termino de las métricas de *dice*, se ve que son del mismo orden entre los conjuntos de entrenamiento, validación y test para las dos bases de datos con valores entre 0,69 y 0,76.

El experimento 10 tiene mejores métricas en la base de datos 1 en la cual ha sido entrenado. En cambio, el experimento 9 que se entrena también sobre la base de datos 1 presenta mejores métricas de *dice* sobre la base de datos 2.

Los experimentos 11 y 12 presentan mejores resultados en la base de datos 2 que ha servido para su entrenamiento. El experimento 14 es el que da mejores resultados de la métrica de *dice* sobre las dos bases de datos.

TABLA 30: RESULTADOS DE LAS MÉTRICAS DE DICE DE LOS EXPERIMENTOS 9 A 14 A NIVEL DE *PATCH*

EXPERIMENTO	*DICE* PROMEDIO						DESVIACIÓN ESTÁNDAR					
	Entrenamiento		Validación		Test		Train		Validación		Test	
BASE DE DATOS	1	2	1	2	1	2	1	2	1	2	1	2
9	0,6988	0,7128	0,7216	0,7315	0,6957	0,7109	0,11	0,11	0,12	0,11	0,11	0,12
10	0,7440	0,7073	0,7497	0,7533	0,7328	0,7262	0,12	0,11	0,12	0,11	0,12	0,12
11	0,7022	0,7507	0,7406	0,7452	0,7283	0,7419	0,12	0,12	0,12	0,11	0,11	0,12
12	0,7232	0,7596	0,7437	0,7597	0,7368	0,7438	0,13	0,12	0,12	0,12	0,12	0,13
13	0,7162	0,7521	0,7450	0,7586	0,7303	0,7473	0,13	0,12	0,12	0,12	0,12	0,13
14	**0,7300**	**0,7553**	**0,7448**	**0,7646**	**0,7358**	**0,7451**	0,12	0,12	0,12	0,12	0,12	0,12

Los experimentos 9; 11 y 13 presentan una arquitectura de red neuronal convolucional de U-Net clásica mientras que los experimentos 10; 12 y 14 tiene una arquitectura de red encoder-decoder con capas residuales añadidas (ResU-Net). La Tabla 31 presenta los resultados según la base de datos en promedio de los experimentos con la arquitectura U-Net y de los experimentos con la arquitectura ResU-Net. Se ve claramente que en todos los conjuntos de datos de las dos bases de datos la métrica de *dice* es más alta con una arquitectura ResU-Net. Esto demuestra de nuevo el interés importante de añadir capas residuales a una red *encoder-decoder* para mejorar los resultados de segmentación.

TABLA 31: COMPARACIÓN DE LAS MÉTRICAS DE RED U-NET V/S RESU-NET A NIVEL DE PATCH

Arquitectura del modelo	*Dice* promedio					
	Entrenamiento		Validación		Test	
	1	2	1	2	1	2
U-Net (experimento 9; 11 y 13)	0,706	0,739	0,736	0,745	0,718	0,733
ResU-Net (experimento 10; 12 y 14)	0,732	0,741	0,746	0,759	0,735	0,738

El modelo del experimento 14 que ha sido entrenado sobre imágenes de las dos bases de datos y que tiene una arquitectura red neuronal convolucional *encoder-decoder* con capas residuales (ResU-Net) a nivel de patch es el mejor modelo obtenido en este trabajo. Sus resultados sobre imágenes que provienen de las dos bases de datos se pueden ver en la Tabla 32. Se puede notar que las métricas a nivel de la imagen reconstruida son ligeramente más altas que a nivel de *patch* debido a la adición de los *patches* sin anotación (máscara solo negra) que son correctos ya que la segmentación gruesa previa no presenta falsos negativos en la mayoría de los casos.

Se visualiza igualmente las métricas después de la etapa de post procesado previamente comentada. Las métricas aumentan de 0.02 en promedio entre las imágenes antes del post procesado y las imágenes después del post procesado. Este incremento es bastante significante para concluir que la etapa de post procesado de la máscara de predicción de la epidermis permite mejorar la segmentación obtenida, sobre todo quitando elementos debidos al ruido.

Se observa también que las métricas son más altas sobre la base de datos CLARIFYv1 que sobre CLARIFYv2, este fenómeno puede ser debido al número mayor de épocas entrenadas sobre las imágenes de la base de datos CLARIFYv1 como al hecho que al ser una base de datos menos amplia

que CLARIFYv2, la base de datos CLARIFYv1 presenta menos variabilidad y puede resultar más fácil de predecir para el modelo.

En la Tabla 32, se añade como última fila los resultados promediados sobre las dos bases de datos. Se obtiene después de la reconstrucción de la máscara de la epidermis al tamaño de la imagen original y después del post procesado un *dice* de 0.78 en el conjunto de entrenamiento, un *dice* de 0.80 en el conjunto de validación y un *dice* de 0.78 en el conjunto de test.

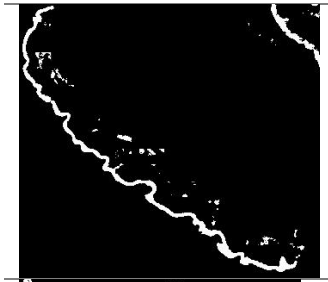
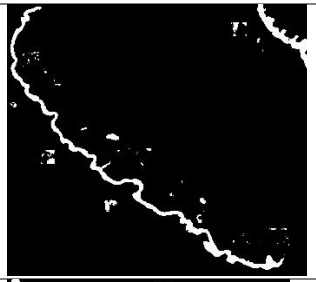
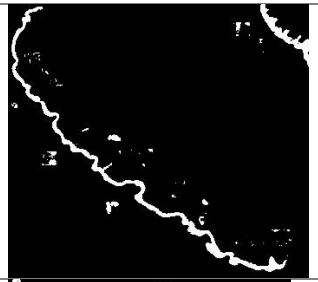
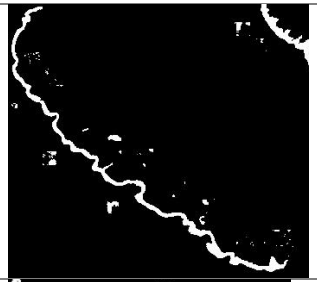
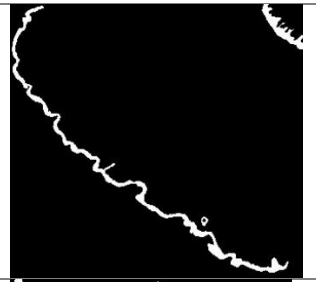
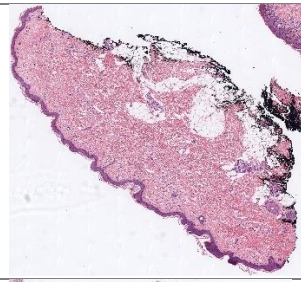
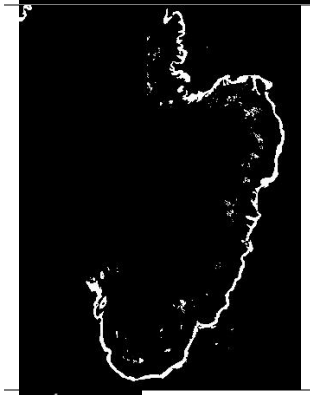
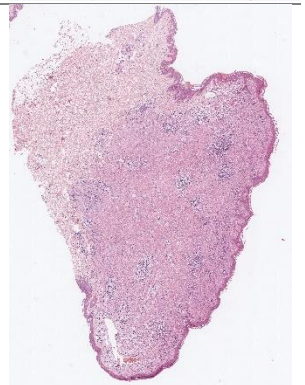
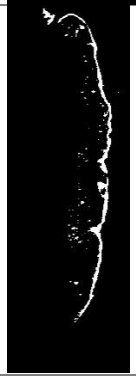
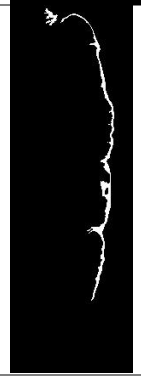
TABLA 32: RESULTADOS DE LAS MÉTRICAS DEL EXPERIMENTO 14 A NIVEL DE PATCH, DE IMAGEN RECONSTRUIDA ANTES Y DESPUÉS DEL POST PROCESADO

BASE DE DATOS	MODELO 14	NIVEL *PATCHES*		NIVEL IMAGEN RECONSTRUIDA		NIVEL IMAGEN RECONSTRUIDA CON POST PROCESADO	
		dice	std	*dice*	std	*dice*	std
CLARIFYV2	Entrenamiento	0,7553	0,12	0,6995	0,10	0,7197	0,10
	Validación	0,7646	0,12	0,7140	0,08	0,7294	0,08
	Test	0,7451	0,12	0,6953	0,09	0,7193	0,09
CLARIFYV1	Entrenamiento	0,7300	0,12	0,7934	0,06	0,8322	0,06
	Validación	0,7448	0,12	0,8406	0,07	0,8738	0,05
	Test	0,7358	0,12	0,8005	0,07	0,8398	0,06
PROMEDIO	Entrenamiento	0,7426	0,12	0,7465	0,08	0,7760	0,08
	Validación	0,7547	0,12	0,7773	0,08	0,8016	0,06
	Test	0,7405	0,12	0,7479	0,08	0,7796	0,07

En la Tabla 33 se presentan resultados de máscaras de predicción reconstruidas de los experimentos 11 a 14. Para conseguir una mejor comparación con los modelos anteriores se usan las mismas imágenes de los conjuntos de test, la primera fila pertenece a la base de datos CLARIFYv2 y las dos filas siguientes a la base de datos CLARIFYv1. Se ven entonces las imágenes con un tamaño reducido pero sin convertirlas a imágenes de 512*512 que como se había visto puede introducir distorsión en la imagen. Las máscaras presentadas en la tabla son en salida de las redes, antes del post procesado y por eso presentan todavía artefactos de ruido.

Las máscaras de segmentación de la epidermis en imágenes histopatológicas de piel con lesiones melanocíticas son parecidas entre los diferentes experimentos. Sin embargo, el modelo 14 presenta las máscaras con menor ruido, por eso sus métricas de dice son más altas y resulta en ser el mejor modelo obtenido en este trabajo. En estos tres casos, las máscaras obtenidas se parecen mucho al *ground truth* y se puede concluir que se obtiene una segmentación satisfaciente de la epidermis.

TABLA 33: EJEMPLO DE MÁSCARAS DE PREDICCIÓN DE LOS EXPERIMENTOS 11 A 14 DE DATOS DE TEST DE LAS DOS BASES DE DATOS

Experimento				Ground truth	Imagen original
11	12	13	14		
					
					
					

5.7. Resumen de los resultados obtenidos y comparación con el estado del arte

La Tabla 34 presenta un resumen de las métricas del modelo 14 después del post procesado en imágenes en promedio sobre las dos bases de datos de imágenes de biopsias de piel con melanoma spitzoide que se disponen para este trabajo. Se obtienen resultados bastante altos que habría ahora que comparar con los resultados que se pueden obtener en la literatura científica sobre el tema. A continuación, se realiza una discusión de los resultados obtenidos con el modelo desarrollado en este trabajo.

TABLA 34: RESUMEN DE LAS MÉTRICAS DEL MEJOR MODELO OBTENIDO DESPUÉS DEL POST PROCESADO

	Mejor modelo: 14	
	Dice	**Desviación estándar**
Entrenamiento	0,7760	0,08
Validación	0,8016	0,06
Test	0,7796	0,07

Los artículos de referencia en segmentación automática de la epidermis en biopsias de piel usan varias métricas y no todos usan la métrica de *dice*, se usa el *Positive Predictive value* (PPV), la sensibilidad y el *Matthew Correlation Coefficient* (MCC) como explicado en la sección 1.3. Entonces la comparación de las métricas entre el trabajo realizado y los artículos del estado del arte es incompleta.

Además de las métricas diferentes, la comparación con los resultados obtenidos en los artículos del estado del arte es una comparativa indirecta porque no se usa la misma base de datos de imágenes. El modelo desarrollado en este trabajo utiliza biopsias de piel con la particularidad de presentar lesiones melanocíticas Spitzoide cuando las bases de datos que usan los artículos de referencia son bases de datos públicas de biopsias de piel sanas, como las imágenes de *University of British Columbia Virtual Slidebox* [53] y de *University of Michigan Virtual Slidebox* [54]. Estas dos bases de datos son de acceso público pero no están anotadas, con lo cual no se han utilizado en este trabajo. El hecho que las imágenes de CLARIFYv1 y CLARIFYv2 presenten lesiones melanocíticas hace que la epidermis sea una zona mucho más difícil de segmentar aun para patólogos experimentados como explicado en la sección 1.2.3.

Encima, no se ha implementado los métodos descritos en los artículos sobre las bases de datos CLARIFYv1 y CLARIFYv2 por falta de detalle sobre el *framework* empleado. Por lo tanto, la comparación expuesta a continuación es una comparativa indirecta y hay que considerarla con el cuidado requerido.

Los resultados obtenidos en este trabajo mejoran los resultados de los algoritmos presentados en el estado del arte. El método presentado en este trabajo tiene una métrica de *dice* en el conjunto de test de 0.78 que supera con una margen importante los resultados de los algoritmos CET ([20]), GTSA ([19]), THM ([22]) y PASC ([23]). Estos métodos tienen como valor del *dice* en datos de test 0.41 para CET, 0.53 para GTSA, 0.42 para THM y 0.59 para PASC. La métrica de *dice* del método obtenido es un 30% mejor que las de los artículos del estado del arte, en promedio.

El método implementado en este trabajo presenta también la ventaja de no presuponer ninguna hipótesis para realizar el análisis de las imágenes, no hay que tener información a priori sobre la intensidad de la tinción o el tamaño de la epidermis. Es un modelo enteramente automático y que realiza la segmentación a dos niveles de precisión de la imagen, mientras manteniendo razonable el coste computacional.

En cambio, el método propuesto en el artículo [24] alcanza un *dice* de 0.86 en datos de test, lo que supera el 0.78 del modelo desarrollado en este trabajo. Pero esta diferencia se debe analizar con cuidado como explicado previamente. Se puede notar que la métrica obtenida en este trabajo se acerca del valor del mejor algoritmo presente en la bibliografía utilizando imágenes con lesiones específicas que complican la segmentación de la epidermis.

Comparativamente con el método del artículo [24] que también hace uso de una arquitectura de red U-Net, el método propuesto en este trabajo filtra los patches que contienen epidermis en la segmentación gruesa para reducir el coste computacional. También se usa como función de pérdida el *dice*, que es más adecuada a la tarea de segmentación de imágenes. Por fin, la adición de capas residuales permite facilitar el entrenamiento de la red *encoder-decoder* y obtener mejores resultados. Las etapas de post procesado del método presentado en este trabajo son mínimas para no introducir sesgo en los resultados y permitir una mejor robustez del modelo. Este modelo ha también podido ser validado con una segunda base de datos lo que permite una validación externa además de la validación interna tradicional.

Entonces el modelo obtenido en este trabajo es el primer que se enfoca a la segmentación de la epidermis únicamente en imágenes con lesiones melanocíticas spitzoides de biopsias de piel. Supera en termino de métricas los algoritmos de segmentación automática con *machine learning*. En cuanto a la otra red convolucional *encoder-decoder* U-Net del artículo [24], las métricas obtenidas son similares pero en este artículo se usan biopsias sanas que no son tan difíciles de segmentar como las que presentan lesiones melanocíticas spitzoides. Además el modelo obtenido en el artículo de 2019 ha podido ser entrenado en bases de datos publicas más amplias, lo que permite obtener métricas más altas.

Para el caso específico de las lesiones melanocíticas y el diagnóstico del melanoma spitzoide este modelo es muy prometedor ya que no existe a la hora de este trabajo modelos similares en la literatura.

Este trabajo demuestra también que la adición de capas residuales a la arquitectura tradicional de la red neuronal convolucional *encoder-decoder* U-Net permite mejorar de manera significativa los resultados obtenidos de cara a la segmentación de la epidermis. Se preconiza entonces usar una arquitectura de ResU-Net para la tarea de segmentación de imágenes histológicas con aprendizaje profundo.

6. Conclusión y líneas futuras

En este proyecto, se ha realizado un estudio de investigación en el ámbito de la patología digital y de los modelos automatizados usando inteligencia artificial. El trabajo se centra en las neoplasias melanocíticas spitzoides que son lesiones cutáneas poco frecuentes y difíciles de diagnosticar. El diagnóstico se hace usando el *gold standard* de la imagen de biopsia de piel y la zona de interés suele ser la epidermis y su junción con la dermis. Actualmente no se ha desarrollado un modelo de segmentación de la epidermis en el contexto específico de estas lesiones que pueden ser un nevus spitzoide benigno o un melanoma spitzoide maligno. El interés de sistemas informáticos de ayuda al diagnóstico para los patólogos en este ámbito es grande.

Una vez documentado el estado del arte de la segmentación de la epidermis en imágenes de biopsias de piel con lesiones melanocíticas y de las redes neuronales convolucionales, se concluye que se debe emplear para este proyecto un modelo compuesto por una red neuronal *encoder-decoder* seguida de otra pero con mayor detalle a nivel de *patch*. La primera red realiza una segmentación gruesa de la zona de la epidermis para delimitar la parte de la imagen en la cual se debe enfocar para lograr una segmentación de precisión de forma computacionalmente más eficaz y rápida que analizando la imagen por completo. La segunda red realiza la segmentación precisa a nivel de patch. Luego se reconstruye la imagen a partir de las predicciones de máscara que localiza la epidermis por completo. Por último, se añade la etapa de post procesado de quitar los artefactos detectados que corresponden a artefactos de ruido y no a la información relevante.

Una vez justificada la elección de las dos arquitecturas del modelo que se quieren implementar y comparar, se procede al análisis y preprocesado de una primera base de datos (CLARIFYv1) que ya lleva sus anotaciones. Para servir como base de datos de validación y de ampliación de la precisión de los modelos implementados, se realiza la anotación manual de una segunda base de datos (CLARIFYv2) bajo la supervisión de un patólogo. Una vez realizadas las anotaciones con el software *MicroDraw*, se extraen las máscaras y los *patches* correspondientes de la WSI. Durante la generación de los *patches* se selecciona automáticamente los *patches* relevantes para el entrenamiento, eliminando aquellos que no contienen tejido a analizar y que estén fuera de la zona tumoral. Se ha realizado la partición aleatoria de los datos en conjuntos de entrenamiento, validación y test con la proporción recomendada y comúnmente usada para estos tipos de modelos de *Deep learning*.

A continuación, con las bases de datos preparadas y las arquitecturas definidas, se han analizado los requerimientos computaciones del proyecto y seleccionado el entorno de programación más adecuado. Se ha empleado el entorno de programación de Pycharm usando un servidor para los datos y otro con una GPU de alta potencia como servidor de computación estando en entorno Docker. Se ha llevado a cabo los entrenamientos de los diferentes modelos para permitir la comparación y la elección del modelo que da los mejores resultados. Se ha elegido adecuadamente las funciones de pérdidas y métricas, así como el optimizador y todos los hiperparámetros necesarios al entrenamiento para obtener un modelo robusto y un entrenamiento relativamente rápido. En la función de pérdida elegida se tiene en cuenta el desbalance de clase usando una versión ponderada de la métrica de *dice*. Es importante tenerlo en cuenta ya que debido a la rareza que presentan las lesiones melanocíticas spitzoides, los tumores difieren considerablemente en tamaño. Debido al bajo número de pacientes involucrados en estas bases de datos, se han realizado otros experimentos con un entrenamiento de los algoritmos usando la técnica de validación cruzada.

El tiempo de entrenamiento es grande pero no supera el día de cálculos así que sigue siendo un modelo bastante fácil de entrenar por si se tienen que afinar los pesos del modelo con nuevos datos disponibles de ampliación de la base de datos de entrenamiento.

Lo que primero destaca de los modelos obtenidos es que la arquitectura con capas residuales permite un modelo que pesa menos (150ko) que la U-Net tradicional (llega casi al Go). Sin embargo, el modelo obtenido permite una segmentación precisa de la máscara de epidermis en imágenes de biopsias de piel en pacientes con melanoma spitzoide. Se ha mostrado en este trabajo que la arquitectura de red neuronal convolucional ResU-Net ofrece un modelo más robusto y preciso con resultados mejores que una red *encoder-decoder* U-Net. Así que el modelo finalmente utilizado es el que conlleva capas residuales.

El modelo automático a dos niveles de resolución obtenido logra una predicción de la máscara de la epidermis con una métrica de *dice* de 0.78 en imágenes de inferencia (conjunto de test) para la arquitectura con la adición de capas residuales. Se obtiene un *dice* de 0.78 en el conjunto de entrenamiento y 0.80 en validación lo que son métricas de predicción altas comparativamente con las obtenidas en la literatura científica del campo.

Se puede concluir que se ha logrado diseñar un sistema automático de segmentación de la epidermis en biopsias con lesiones melanocíticas spitzoides. Los resultados obtenidos en este trabajo representan un avance para el diagnóstico con una mayor digitalización y automatización de los tumores melanocíticos spitzoides.

Las limitaciones de la solución propuesta son el coste computacional que sigue importante y la calidad de ciertas anotaciones utilizadas de la epidermis. Con una mayor cantidad de datos de entrenamiento se podría lograr mejorar estos resultados. No obstante, la adquisición de nuevas imágenes histopatológicas es un proceso costoso que requiere bastante tiempo. Además, este trabajo se ha llevado a cabo solo con imágenes de una misma fuente (mismo hospital, periodo y proceso de obtención de las biopsias). Por lo que seguramente las imágenes utilizadas cuenten con cierto sesgo. Para demostrar la robustez del modelo habría que validarlo en datos de otros hospitales. El número de muestras utilizado es relativamente bajo en este proyecto y por eso se necesitara en un futuro una validación externa que constituye la línea futura de trabajo principal .

Como líneas futuras de trabajo, se propone integrar este modelo dentro de un programa que detecta regiones de potenciales lesiones melanocíticas para ayudar el patólogo en su diagnóstico complejo del melanoma Spitzoide. El futuro de este proyecto se relaciona con el del proyecto europeo CLARIFY en el cual se enmarca. Una línea de trabajo que tiene el proyecto y que se tiene que considerar es la implementación de un sistema como el diseñado aquí pero en la nube para facilitar la interpretación de las imágenes de biopsias. Para ello, se podría desarrollar una plataforma web que integre los modelos obtenidos en este estudio y que permita la segmentación de la epidermis de nuevas imágenes subidas a la nube por el patólogo. Permitiría servir de sistema de apoyo al diagnóstico de los patólogos independientemente de su localización.

7. Referencias bibliográficas

1. GUY, G. P., et al. (2015). Vital signs: melanoma incidence and mortality trends and projections - United States, 1982-2030. *MMWR. Morbidity and mortality weekly report*, *64*(21), 591–596.
2. ORGANIZACIÓN MUNDIAL DE LA SALUD. *Radiation: Ultraviolet (UV) radiation and skin cancer*. <https://www.who.int/news-room/questions-and-answers/item/radiation-ultraviolet-(uv)-radiation-and-skin-cancer> [Consulta: 12/04/2022]
3. WEIR, H., et al. (2011). Melanoma in adolescents and young adults (ages 15-39 years): United States, 1999-2006. *J Am Acad Dermatol*. Estados Unidos, 65(5 Suppl 1):S38-49.
4. KAMINO H. (2009). Spitzoid melanoma. *Clinics in dermatology*, *27*(6), 545–555. https://doi.org/10.1016/j.clindermatol.2008.09.013
5. ELMORE, J. G., et al. (2017). Pathologists' diagnosis of invasive melanoma and melanocytic proliferations: observer accuracy and reproducibility study. *BMJ (Clinical research ed.)*, *357*, j2813. https://doi.org/10.1136/bmj.j2813
6. E-CANCER. *La peau*. <https://www.e-cancer.fr/Patients-et-proches/Les-cancers/Melanome-de-la-peau/La-peau> [Consulta: 12/04/2022]
7. MEDLINE PLUS. *Capas de la piel*. <https://medlineplus.gov/spanish/ency/esp_imagepages/8912.htm> [Consulta: 09/05/2022]
8. HERNANDEZ, A. La célula, unidad estructural y funcional. <https://es.slideshare.net/Alberkar/la-clula-unidad-estructural-y-funcional-el-ncleo-2013-81584011> [Consulta: 12/04/2022]
9. BERA, K., et al. (2019). Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. *Nature reviews. Clinical oncology*, *16*(11), 703–715. https://doi.org/10.1038/s41571-019-0252-y
10. MEDICINA ESTESTICA DR. CARMEN TRASEIRA. *Hidratación fisiológica de la piel*. <https://www.medicinaesteticactraseira.es/hidratacion-fisiologica-de-la-piel/> [Consulta: 06/05/2022]
11. GARWAL, S. & KRISHNAMURTHY, K. (2022). "Histology Skin" en *StatPearls*. Estados Unidos, accesible a https://www.ncbi.nlm.nih.gov/books/NBK537325/
12. MEGIAS, M. et al. (2019). *Altas de histología vegetal y animal*. España, accesible a http://mmegias.webs.uvigo.es/inicio.html
13. MACEO, A. (2003). *Anatomía Y Fisiología De La Cresta De Fricción En La Piel Adulta*. Capítulo 2
14. SCHNEIDER, M. R., SCHMIDT-ULLRICH, R. & PAUS, R. (2009). The hair follicle as a dynamic miniorgan. *Current biology : CB*, *19*(3), R132–R142. https://doi.org/10.1016/j.cub.2008.12.005

15. HC MARBELLA. *ABCDE del melanoma*. <https://www.hcmarbella.com/es/abcde-del-melanoma/> [Consulta: 09/05/2022]
16. SAINZ-GASPAR, L. et al. (2020). Spitz Nevus and Other Spitzoid Tumors in Children —Part 1: Clinical, Histopathologic, and Immunohistochemical Features, Actas Dermo-Sifiliográficas (English Edition), Volume 111, Issue 1, Pages 7-19, ISSN 1578-2190, https://doi.org/10.1016/j.adengl.2019.12.006.
17. SLIDES SHARE. *Repaso patología sistémica: Piel*. <https://pt.slideshare.net/jihansimonhasbun1/piel-23378636> [Consulta: 06/05/2022]
18. CLARIFY. Cloud Artificial Intelligence for pathology. <http://www.clarify-project.eu/> [Consulta: 12/04/2022]
19. LU, C. & MANDAL, M. (2012). Automated segmentation and analysis of the epidermis area in skin histopathological images. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, *2012*, 5355–5359. https://doi.org/10.1109/EMBC.2012.6347204
20. HAGGERTY, J.M., *et al.* (2014). Segmentation of epidermal tissue with histopathological damage in images of haematoxylin and eosin stained human skin. *BMC Med Imaging* **14,** 7 https://doi.org/10.1186/1471-2342-14-7
21. MOKHTARI, M., et al. (2014). Computer aided measurement of melanoma depth of invasion in microscopic images. *Micron (Oxford, England : 1993)*, *61*, 40–48. https://doi.org/10.1016/j.micron.2014.02.001
22. XU, H. & MANDAL, M. (2015). Epidermis segmentation in skin histopathological images based on thickness measurement and k-means algorithm. *J Image Video Proc.* **2015,** 18 https://doi.org/10.1186/s13640-015-0076-3
23. KLECZEK, P. et al. (2017). Automated epidermis segmentation in histopathological images of human skin stained with hematoxylin and eosin. Proc. SPIE 10140, Medical Imaging 2017: Digital Pathology, https://doi.org/10.1117/12.2249018
24. OSKAL, K.R.J., *et al.* A U-net based approach to epidermal tissue segmentation in whole slide histopathological images. *SN Appl. Sci.* **1,** 672 (2019). https://doi.org/10.1007/s42452-019-0694-y
25. OTSU, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics.* Japon, vol. 9, no. 1, pp. 62-66
26. GU, F., et al. (2018). Multi-Resolution Networks for Semantic Segmentation in Whole Slide Images. *MICCAI COMPAY 2018 Workshop*. Suecia, arXiv:1807.09607v1
27. MALATHI, M. & SINTHIA, P. (2019). Brain Tumour Segmentation Using Convolutional Neural Network with Tensor Flow. Asian Pacific Journal of Cancer Prevention : APJCP. 20(7):2095-2101. DOI: 10.31557/apjcp.2019.20.7.2095. PMID: 31350971; PMCID: PMC6745230.

28. WANG, S., et al. (2019). Pathology Image Analysis Using Segmentation Deep Learning Algorithms. *The American journal of pathology*, *189*(9), 1686–1698. https://doi.org/10.1016/j.ajpath.2019.05.007
29. SLIDE PLAYER. *La chimie de la perception*. <https://slideplayer.fr/slide/1315285/> [Consulta: 06/05/2022]
30. WANG, Y. et al. (2020). The Influence of the Activation Function in a Convolution Neural Network Model of Facial Expression Recognition. *Applied Sciences*, 10(5):1897. https://doi.org/10.3390/app10051897
31. DEEP AI. *Perceptron*. <https://deepai.org/machine-learning-glossary-and-terms/perceptron> [Consulta: 06/05/2022]
32. DUMOULIN, V. & VISIN, F. (2016). A guide to convolution arithmetic for deep learning. Arxiv. Francia, 1603:07285
33. GITHUB. *Convolution steps*. <https://cs231n.github.io/assets/conv-demo/index.html > [Consulta: 21/04/2022]
34. AGARAP, F. (2018). Deep Learning using Rectified Linear Units. A*rXiv*. Filipinas, 1803.08375 https://arxiv.org/abs/1803.08375
35. ANALYTICS VIDHYA. *What is the Convolutional Neural Network Architecture?* <https://www.analyticsvidhya.com/blog/2020/10/what-is-the-convolutional-neural-network-architecture/> [Consulta: 09/05/2022]
36. ARMONI A. (1998). Use of neural networks in medical diagnosis. *M.D. computing : computers in medical practice*, *15*(2), 100–104.
37. YAMASHITA, R., *et al.* (2018). Convolutional neural networks: an overview and application in radiology. *Insights Imaging* **9,** 611–629. https://doi.org/10.1007/s13244-018-0639-9
38. SULTANA, F., SUFIAN, A., & DUTTA, P. (2020). Evolution of Image Segmentation using Deep Convolutional Neural Network: A Survey. *ArXiv, abs/2001.04074*.
39. LIU, K. *et al*. (2021). "Learning Melanocytic Proliferation Segmentation in Histopathology Images from Imperfect Annotations," *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 3761-3770, doi: 10.1109/CVPRW53098.2021.00417.
40. KUCHARSKI, D. et al. (2020). Semi-Supervised Nests of Melanocytes Segmentation Method Using Convolutional Autoencoders. *Sensors (Basel, Switzerland)*, *20*(6), 1546. https://doi.org/10.3390/s20061546
41. RONNEBERGER, O., FISHER, P., BROX, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

42. ZHENGXIN, Z., QINGJIE, L. & WANG, Y. (2018). "Road Extraction by Deep Residual U-Net" en *IEEE Geoscience and Remote Sensing Letters*. China, vol. 15, no. 5, pp. 749-753, doi: 10.1109/LGRS.2018.2802944
43. HE, K., ZHANG, X., REN, S., & SUN, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778.
44. KALAPAHAR, A., et al. (2020). Gleason Grading of Histology Prostate Images Through Semantic Segmentation via Residual U-Net. *2020 IEEE International Conference on Image Processing (ICIP)*, 2501-2505.
45. PASTEUR. *MicroDraw*. < https://microdraw.pasteur.fr/> [Consulta: 16/05/2022]
46. MACENKO, M., et al. (2009). A method for normalizing histology slides for quantitative analysis. *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 1107-1110.
47. IOFFE, S. y SZEGEDY, C. (2015). Batch normalization: accelerating Deep network training by reducing internal covariate shift. In Proceedings of the *32nd International Conference on Machine Learning*. Volume 37, 448-456.
48. BERTELS, J., et al. (2019). Optimizing the Dice Score and Jaccard Index for Medical Image Segmentation: Theory and Practice. *LNCS 11765, Springer Nature Switzerland AG*. China, arXiv:1911.01685
49. TOWARDS DATA SCIENCE. *Metrics to evaluate your semantic segmentation model*. <https://towardsdatascience.com/metrics-to-evaluate-your-semantic-segmentation-model-6bcb99639aa2> [Consulta: 11/05/2022]
50. KINGMA, D.P. & BA, J. (2015). Adam: A Method for Stochastic Optimization. *CoRR, abs/1412.6980*.
51. YING, X. (2019). "An Overview of Overfitting and its solutions" en *Journal of Physics*. China, 1168
52. GURPREET, B. & DEEPAK, G. (2011). An Enhanced Approach to improve the contrast of Images having bad light by Detecting and Extracting their Background. International Journal of Computer Science and Management Studies. 11.
53. THE UNIVERSITY OF BRITISH COLOMBIA. *Virtual collections*. <https://pathology.ubc.ca/virtual-collections/> [Consulta: 24/05/2022]
54. UNIVERSITY OF MICHIGAN. Virtual Slide box. <https://www.pathology.med.umich.edu/apps/slides/> [Consulta: 24/05/2022]

II. Presupuesto

En esta sección se va a proponer un potencial presupuesto de la realización de un modelo automático de segmentación de la epidermis mediante redes encoder-decoder residuales para imágenes histopatológicas de melanoma Spitzoide.

A. Costes de personal

En el master de ingeniería biomédica de la UPV, el trabajo de fin de master cuenta para 20 créditos ECTS lo que corresponde aproximadamente a 500 horas. Se ha dividido estas horas dedicadas a este trabajo según los objetivos mayores que son realizar el estado del arte, preparar la base de datos para su entrada en el modelo, implementar las redes convolucionales residuales para entrenarlas y predecir nuevas entradas. Se ha dedicado aproximadamente 60 horas a la redacción de la memoria de este trabajo. El detalle del cálculo de los costes de personal de este trabajo esta presentado en la Tabla 35.

TABLA 35: DETALLE DE LOS COSTES DE PERSONAL

Descripción de la actividad	Unidad	Medición	Precio	Importe
Estado del arte	h	40	6€/h	240€
Preparación de la base de datos	h	70	13€/h	910€
Implementación de las redes convolucionales residuales y nuevas predicciones	h	330	13€/h	4 290€
Elaboración de la memoria	h	60	6€/h	360€
Costes directos				5 800€
Costes directos complementarios (2%)				116€
Total				5 916€

B. Costes de material hardware

Como material hardware utilizado en este trabajo ya se han mencionado un ordenador con un procesador Intel Core i7-1165G7 de 12 MB de caché hasta 4,7GHz y un sistema operativo Windows 10 de 64 bits, el servidor de computación con procesador Intel i7 @4.20GHz y una tarjeta gráfica NVIDIA Titan XP y la NAS Synology DS918. La NAS vale aproximadamente 750€ y el ordenador 800€. El procesador del servidor de computación vale aproximadamente 345€ y la tarjeta gráfica 1349€, lo que hace un coste del servidor de computación de 1694€. El servidor de computación y de almacenamiento (NAS) son proporcionados por el CVBLab mientras que el ordenador es una propiedad personal del estudiante.

Además para la anotación de las imágenes de la base de datos CLARIFYv2 se ha utilizado una Tablet *Xiaomi Pad* 5 de 6 GB de RAM y un disco de 128 GB para 399€ con su *Smart pen* para 70€ para un coste global de la Tablet de 469€.

El detalle del cálculo de los costes de material hardware de este trabajo esta presentado en la Tabla 36. Se ha aproximado el periodo de tiempo que se pueden usar estos dispositivos y el intervalo que se han visto usado para el desarrollo de este trabajo.

TABLA 36: DETALLE DE LOS COSTES DE MATERIAL HARDWARE

Descripción	Cantidad	Coste unitario en €	Periodo de amortización (meses)	Intervalo amortizado	Coste imputable en €
DELL Inspiron 14500 Intel Core i7-1165G7	1 u	800	72	8	88,89
Servidor de computación	1 u	1694	48	3	108,87
NAS Synology DS918	1 u	750	48	3	46,89
Tablet Xiaomi Pad 5	1 u	469	12	2	78,16
Costes directos					319,81
Costes directos complementarios (2%)					6,39
Total					326,21

C. Costes de material software

El software usado es mayoritariamente *open source* como todas las librerías de Python, la plataforma *Pycharm* y la interfaz *MobaXterm*. En cambio, la UPV provee una versión profesional de *Pycharm* que es de uso pagante y una licencia MatLab. La licencia para un año de Microsoft Office para estudiantes cuesta 149€ al año. El detalle del cálculo de los costes de material software de este trabajo esta presentado en la Tabla 37.

TABLA 37: DETALLE DE LOS COSTES DE MATERIAL SOFTWARE

Descripción	Cantidad	Coste unitario en €	Periodo de amortización (meses)	Intervalo amortizado	Coste imputable en €
Pycharm	1 u	199	12	8	132,67
MatLab	1 u	800	12	8	533,33
Microsoft Office	1 u	149	12	8	99,33
Total					765,34

D. Presupuesto total

Luego, se calcula el importe total del desarrollo de este Trabajo de Fin de Master añadiendo los costes generales, el beneficio industrial y el impuesto sobre el valor añadido (IVA). Se considera un 13% correspondiente a los gastos generales, un 6% al beneficio industrial y un 21% de IVA. El detalle del cálculo del presupuesto total de este trabajo esta presentado en la Tabla 38.

TABLA 38: PRESUPUESTO TOTAL

Descripción	Coste imputable en €
Costes de personal	5 916
Costes de material hardware	326,21
Costes de material software	765,34
Suma	7007,55
Beneficio industrial (6%)	420,45
Costes generales (13%)	910,98
IVA (21%)	1 471,59
Total	9 810,57

Por tanto, el coste total de la realización de este proyecto asciende a **NUEVE MIL OCHOCIENTOS DIEZ CON CINCUENTA Y SIETE CENTIMOS.**