



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



ESCUELA TÉCNICA  
SUPERIOR INGENIEROS  
INDUSTRIALES VALENCIA

**TRABAJO FIN DE GRADO EN INGENIERÍA BIOMÉDICA**

**DISEÑO Y DESARROLLO DE SISTEMAS  
AUTOMÁTICOS DE SEGMENTACIÓN  
APLICADOS A IMÁGENES OCT DE  
RETINA Y PIEL**

AUTORA: M<sup>a</sup> Rocío del Amor del Amor

TUTORA: Valery Naranjo Ornedo

COTUTOR: Adrián Colomer Granero

**Curso Académico: 2018-2019**



# Agradecimientos

“ A Valery, por darme la oportunidad de adentrarme en este maravilloso mundo,  
por todo lo que he aprendido, tanto en lo personal como en lo profesional,  
y por toda su confianza depositada en mí.

A Adri, por su apoyo incondicional  
y por hacer que este trabajo fuera posible.

A mi familia, por ser mi mayor pilar,  
por quererme tal y como soy,  
y por todo el apoyo que he recibido.

A mis amigos, en especial a mis cariiis,  
por todo el amor durante estos cuatro años.

A todo el grupo de CVBlab, por ser tan increíbles,  
y por haber encontrado el verdadero compañerismo.

A mi Ricis, por no creer en las casualidades, por ser tan coherente,  
por el cariño y por hacer que siempre sea la penúltima vez.

A todos, muchas gracias.”



# Resumen

En el presente Trabajo Fin de Grado (TFG) se pretende abordar el desarrollo de algoritmos automáticos de segmentación basados en *deep learning* con los que poder identificar tanto las diferentes capas retinianas como las principales capas de la piel visibles con Tomografía de Coherencia Óptica (OCT). El objetivo final reside en la obtención de las estructuras que son discriminantes a la hora de realizar el diagnóstico de enfermedades que afectan tanto a la retina, glaucoma, como a la piel, cáncer de piel y psoriasis.

Para conseguir el objetivo propuesto, se han utilizado arquitecturas enconder-decoder basadas en redes neuronales convolucionales. La arquitectura propuesta para la segmentación de retina es muy similar a la planteada para piel, con la salvedad de que esta última presenta una mayor profundidad dado que este tipo de imágenes son más complejas y requieren la obtención de un mayor número de características. Para el desarrollo del algoritmo de segmentación de piel, se ha contado con 270 imágenes pertenecientes a 10 pacientes sanos. En el caso de la retina, se han utilizado un total de 129 imágenes de 10 roedores diferentes a los que se le ha inyectado ET-1, un potente vasoconstrictor que permite simular la presencia de glaucoma.

Para llevar a cabo el entrenamiento de dichos algoritmos, dado que estos son supervisados, es necesario generar las etiquetas correspondientes, es decir, realizar una segmentación manual. Para ello, se ha implementado una aplicación que permite realizar segmentaciones manuales o modificarlas en el caso de que ya existan. A partir de las muestras disponibles, se ha desarrollado una partición de datos que garantiza que los modelos generados tras el entrenamiento de los algoritmos son robustos y proporcionan unos resultados fiables. Tras el entrenamiento y testeo de los algoritmos propuestos, se ha realizado un robusto proceso de post-procesado. Este ha sido diferente en piel y en retina puesto que los objetivos para los que ha sido diseñado en cada caso son completamente diferentes.

Los resultados obtenidos, tras aplicar los métodos automáticos propuestos, han sido comparados con la segmentación manual previamente realizada con el fin de computar diferentes métricas que evalúen la bondad de los algoritmos. Además, se ha realizado una comparación objetiva de los resultados adquiridos con los publicados en otros estudios del estado del arte.

**Palabras Clave:** tomografía de coherencia óptica, segmentación semántica, aprendizaje profundo, red neuronal convolucional, folículos pilosos, capas de la retina.



# Resum

En el present Treball Fi de Grau (TFG) es pretén abordar el desenrotllament d'algortmes automàtics de segmentació basats en extitdeep learning amb els que poder identificar tant les diferents capes retinianas com les principals capes de la pell visibles amb Tomografia de Coherència Òptica (OCT) . L'objectiu final residix en l'obtenció de les estructures que són discriminants a l'hora de realitzar el diagnòstic de malalties que afecten tant la retina, glaucoma, com a la pell, càncer de pell i psoriasi.

Per a aconseguir l'objectiu proposat, s'han utilitzat arquitectures enconder-decoder basades en xarxes neuronals convolucionales. L'arquitectura proposada per a la segmentació de retina és molt semblant a la plantejada per a pell, amb l'excepció que esta última presenta una més profunditat atés que este tipus d'imatges són més complexes i requerixen l'obtenció d'un nombre més gran de característiques. Per al desenrotllament de l'algortme de segmentació de pell, s'ha comptat amb 270 imatges pertanyents a 10 pacients sans. En el cas de la retina, s'han utilitzat un total de 129 imatges de 10 rosegadors diferents dels que se li ha injectat ET-1, un potent vasoconstrictor que permet simular la presència de glaucoma.

Per a dur a terme l'entrenament dels dits algortmes, atés que estos són supervisats, és necessari generar les etiquetes corresponents, és a dir, realitzar una segmentació manual. Per a això, s'ha desenrotllat una aplicació que permet realitzar segmentacions manuals o modificar-les en el cas que ja existisquen. A partir de les mostres disponibles, s'ha desenrotllat una partició de dades que garantix que els models generats després de l'entrenament són robustos i proporcionen uns resultats fiables. Després de l'entrenament i testeo dels algortmes proposats, s'ha realitzat un robust procés de post-processat. Este ha sigut diferent en pell i en retina ja que els objectius per als que ha sigut dissenyat en cada cas són completament diferents.

Els resultats obtinguts, després d'aplicar els mètodes automàtics proposats, han sigut comparats amb la segmentació manual prèviament realitzada a fi de computar diferents mètriques que avaluen la bondat dels algortmes. A més, s'ha realitzat una comparació objectiva dels resultats adquirits amb els publicats en altres estudis de l'estat de l'art.

**Palabras Clave:** tomografia de coherència òptica, segmentació semàntica, aprenentatge profund, xarxa neuronal convolucional, follícles pilosos, capes de la retina.





# Abstract

This Final Degree Project (TFG) aims to address the development of automatic segmentation algorithms based on textit deep learning with which to identify both the different retinal layers and the main layers of the skin visible with Coherence Tomography Optics (OCT). The final objective is to obtain structures that are discriminant when it comes to diagnosing diseases that affect both the retina, glaucoma, skin, skin cancer and psoriasis.

To achieve the proposed objective, encoder-decoder architectures based on convolutional neural networks have been used. The proposed architecture for retinal segmentation is very similar to the one proposed for skin, with the exception that the latter has a greater depth since this type of images are more complex and require obtaining a greater number of characteristics. For the development of the skin segmentation algorithm, 270 images belonging to 10 healthy patients were available. In the case of the retina, a total of 129 images of 10 different rodents that have been injected with ET-1, a powerful vasoconstrictor that simulates the presence of glaucoma, have been used.

To carry out the training of said algorithms, since these are supervised, it is necessary to generate the corresponding labels, that is, perform a manual segmentation. To do this, an application has been developed that allows manual segmentation or modify them if they already exist. From the available samples, a data partition has been developed that guarantees that the models generated after the training are robust and provide reliable results. After the training and testing of the proposed algorithms, a robust post-processing process has been carried out. This has been different in skin and retina since the objectives for which it has been designed in each case are completely different.

The results obtained, after applying the proposed automatic methods, have been compared with the manual segmentation previously performed in order to compute different metrics that evaluate the goodness of the algorithms. In addition, an objective comparison has been made of the results obtained with those published in other state of the art studies.

**Keywords:** optical coherence tomography, semantic segmentation, deep learning, convolutional neuronal network, hair follicles, layers of the retina.



# Listado de abreviaturas principales

OCT	Optical Coherence Tomography (Tomografía de coherencia óptica)
RNFL	Retinal Nerve Fiber Layer (Capa de fibras del nervio óptico)
GCL	Ganglion Cell Layer (Capa de células ganglionales)
IPL	Inner Plexiform Layer (Capa plexiforme interna)
IS-OS	Inner Segment and Outer Segment (Segmento interno y externo)
INL	Inner Nuclear Layer (Membrana nuclear interna)
IML	Inner Limiting Membrane (Membrana limitante interna)
CNN	Convolutional Neural Network (Red neuronal convolucional).
SGD	Stochastic Gradient Descent (Gradiente de descenso estocástico)



# Índice general

<b>I Memoria</b>	<b>1</b>
<b>1 Introducción</b>	<b>3</b>
1.1 Motivación y descripción del problema . . . . .	4
1.2 Marco del proyecto . . . . .	5
1.3 Objetivos del proyecto . . . . .	5
1.4 Estructura del documento . . . . .	6
<b>2 Marco teórico</b>	<b>7</b>
2.1 Ojo . . . . .	8
2.2 Piel . . . . .	11
2.3 Tomografía de coherencia óptica . . . . .	14
<b>3 Materiales y métodos</b>	<b>23</b>
3.1 Materiales . . . . .	24
3.2 Metodología . . . . .	28
<b>4 Resultados</b>	<b>59</b>
4.1 Resultados obtenidos en imágenes retinianas . . . . .	61
4.2 Resultados obtenidos en imágenes epiteliales . . . . .	65
<b>5 Conclusiones y líneas futuras</b>	<b>69</b>
5.1 Conclusiones . . . . .	70
5.2 Líneas futuras . . . . .	71
<b>Bibliografía</b>	<b>73</b>

II Presupuesto.	79
1 Presupuesto	81
1.1 Objetivo . . . . .	82
1.2 Presupuestos parciales . . . . .	82
1.3 Presupuesto Total . . . . .	84

**Parte I**

**Memoria**





# Capítulo 1

## Introducción

### *Índice de contenidos*

1.1 Motivación y descripción del problema . . . . .	4
1.2 Marco del proyecto . . . . .	5
1.3 Objetivos del proyecto . . . . .	5
1.4 Estructura del documento . . . . .	6

## 1.1 Motivación y descripción del problema

Según la Organización Mundial de la Salud (OMS), el glaucoma es una de las principales causas de ceguera. Aproximadamente 5,2 millones de personas ciegas lo son debido al glaucoma, lo que supone un 15 % del total. Además, se calcula que hay unos 60 millones de afectados y, debido al envejecimiento de la población, estas cifras podían incrementarse a 111,8 millones para el año 2040 [1]. La detección temprana de esta enfermedad es importante ya que, aunque el daño no puede ser reparado, la degradación puede detenerse si esta afección es tratada con láser o con gotas para los ojos durante sus primeras etapas.

Por otro lado, las enfermedades que afectan a la piel también tienen relevancia en la población, pues el cáncer de piel es el segundo tipo de cáncer más común en el mundo [2]. Solo en España, en los últimos cuatro años, se ha producido un aumento del 38 % y, además, se prevé que esta tendencia continúe en las siguientes décadas [3]. Otra enfermedad dermatológica importante es la psoriasis. Esta afecta al 2 % de la población y en personas mayores de 65 años puede derivar en complicaciones tales como el desarrollo de linfomas. [4]. En las enfermedades dermatológicas el diagnóstico precoz también resulta fundamental para la mejor actuación del tratamiento.

Las técnicas de diagnóstico invasivas como las biopsias permiten analizar con máximo detalle las estructuras internas y llevar a cabo un diagnóstico precoz. Sin embargo, en algunos casos causan dolor y daños en los tejidos adyacentes [5]. En la última década, los avances en óptica, fibra y tecnología láser, han permitido el desarrollo de técnicas de diagnóstico novedosas. Entre ellas, destaca la tomografía de coherencia óptica (OCT), una técnica de diagnóstico por imagen no invasiva. La OCT es una tecnología interferométrica que generalmente utiliza luz infrarroja, lo que permite una penetración profunda [6]. Este hecho posibilita que, mediante esta técnica, se puedan obtener los mismos detalles que en una biopsia, con la ventaja de que el cuerpo humano no se ve dañado.

La OCT se utiliza en gran medida para la detección de enfermedades oculares y, actualmente, está ganando relevancia en el diagnóstico de enfermedades epiteliales, ya que permite alcanzar una profundidad de hasta 2 mm, llegando a visualizar las estructuras más importantes de la piel [5].

Muchas de las enfermedades que afectan al ojo producen cambios en la morfología de la retina, comprometiendo el grosor de las capas más importantes que la conforman. Por otro lado, en las enfermedades que afectan a la piel, la epidermis resulta la capa más afectada. La obtención de estas estructuras en las imágenes de OCT, por parte de los especialistas para su posterior análisis, resulta fundamental para llevar a cabo un diagnóstico precoz. Es por ello, que una segmentación automática de estas capas podría reducir la carga de trabajo de los expertos en anatomía retiniana y epitelial, ayudándoles a mejorar en términos de coste-efectividad.

Por todo lo descrito anteriormente, este trabajo pretende desarrollar algoritmos capaces de segmentar automáticamente las estructuras más relevantes, tanto en imágenes OCT de retina como de piel, que sean robustos y generalizables. Por consiguiente, con la creación de estos algoritmos de segmentación se pretende proporcionar un sistema de ayuda al diagnóstico que facilite la tarea de los especialistas a la hora de diagnosticar enfermedades asociadas a la retina y a la piel.

## 1.2 Marco del proyecto

Este trabajo se enmarca dentro del proyecto de investigación “*Glaucoma – Advanced, Label-free High resolution Automated OCT Diagnostics*” (GALAHAD). GALAHAD es un proyecto europeo coordinado por *Gooch Housego (Torquay) Ltd.* en el que participa el grupo de investigación CVBLab (Computer Vision and Behaviour Analysis Lab) perteneciente a la Universidad Politécnica de Valencia. Este proyecto se encuentra subvencionado por Horizonte2020, programa de Financiación de la Investigación y la Innovación en la Unión Europea, para el periodo 2016-2020.

El propósito general de este proyecto se centra en la necesidad de mejorar los sistemas de diagnóstico por imagen para conseguir una detección precoz del glaucoma. El cribado actual y el diagnóstico básico de esta enfermedad, tales como la medición de la presión intraocular, las pruebas de campo visual y la detección de daños estructurales en la cabeza del nervio óptico y en la capa de la fibra nerviosa de la retina, presentan altas tasas de falsos positivos y/o falsos negativos. Esto se debe a que no es posible realizar un análisis en profundidad de los daños en la cabeza del nervio óptico debido a la mala resolución y precisión de las tecnologías disponibles. El candidato principal para lidiar con este problema es la tomografía de coherencia óptica. Sin embargo, para llevar a cabo un buen diagnóstico, la resolución axial requerida es de  $1 \mu\text{m}$ , mucho más allá de la resolución de  $3\text{-}5 \mu\text{m}$  que presentan los sistemas comerciales actuales. Es por ello que, GALAHAD pretende desarrollar un sistema de diagnóstico de OCT de alta resolución, libre de etiquetas, compacto y de fácil manejo. Una vez desarrollado, dada su alta resolución, este podrá aplicarse de forma satisfactoria a otros órganos de estudio como la piel.

Dado que dentro del marco de este proyecto, la labor de la Universidad Politécnica de Valencia (UPV), más concretamente del grupo CVBLab, es el diseño y desarrollo de algoritmos automáticos para que sean aplicados sobre las imágenes OCT obtenidas, este trabajo está centrado en la confección de métodos destinados a la segmentación, una de las labores más importantes a la hora de efectuar un buen diagnóstico.

## 1.3 Objetivos del proyecto

El objetivo principal de este trabajo fin de grado es el desarrollo de algoritmos automáticos basados en redes neuronales convoluciones (CNN) para la segmentación de las estructuras relevantes de la piel y del ojo que son visibles en una imagen de OCT y que permiten la detección de enfermedades. Para lograr la consecución del objetivo principal de este TFG se debe abordar una serie de objetivos específicos:

1. Obtener las imágenes que son objeto de estudio y adecuarlas para que puedan ser utilizadas.
2. Realizar un estado del arte de los métodos que anterior y actualmente se están utilizando para suplir los objetivos de segmentación planteados en este proyecto.
3. Implementar una aplicación que permita realizar y/o modificar manualmente segmentaciones en las imágenes de interés.
4. Generar un *ground truth*<sup>1</sup> de la segmentación de las estructuras objetivo tanto para la base de datos de piel como para la de retina. Esto es imprescindible puesto que se va a trabajar con algoritmos de segmentación supervisados. Así el *ground truth* se usará tanto para entrenar los modelos durante la etapa de entrenamiento, como para validarlos durante la fase de test.

---

<sup>1</sup>Ground truth: Segmentación manual utilizada como referencia con la que se comparan los resultados obtenidos por un método automático.

5. Llevar a cabo una partición de datos que garantice la robustez del modelo al testear con imágenes diferentes a las que se ha entrenado. Dicha partición no debe ser aleatoria, sino que se ha de desarrollar para que las imágenes correspondientes a un paciente nunca se utilicen para entrenar y testear en la misma iteración, asegurando así la fiabilidad del método.
6. Realizar un post-procesado en los mapas de segmentación obtenidos a la salida de la red neuronal convolucional, con la finalidad de mejorar los resultados.
7. Efectuar una comparación entre los diferentes métodos de segmentación propuestos en este trabajo y los propuestos por otros autores del estado del arte actual en la medida que sea posible.
8. Identificar y describir los problemas y limitaciones encontrados durante el estudio y, proponer posibles mejoras para investigaciones futuras.

## **1.4 Estructura del documento**

En el capítulo 2 se realiza una descripción anatómica del ojo, profundizando en la retina, y de la piel puesto que son los órganos objetivo del presente trabajo. Además, se exponen las características fundamentales de la OCT, así como su utilidad clínica en las enfermedades relativas al ojo y a la piel.

En el capítulo 3 se hace referencia al material y a la metodología empleada para llevar a cabo el desarrollo de los algoritmos de segmentación en retina y piel.

En el capítulo 4 se muestran, se discuten y se comparan con el estado del arte los resultados obtenidos con las técnicas propuestas para la segmentación de imágenes. Además, se aportan los resultados obtenidos tras el análisis de nuevas imágenes de retina totalmente desconocidas por el sistema, tal y como se llevaría a cabo para nuevas muestras en la práctica clínica.

En el capítulo 5 se resumen las conclusiones más importantes extraídas de los estudios realizados y se detallan algunas propuestas, así como posibles líneas de investigación dentro del marco del proyecto.

## Capítulo 2

# Marco teórico

### *Índice de contenidos*

2.1 Ojo . . . . .	8
2.1.1 Anatomía del ojo . . . . .	8
2.1.2 Glaucoma como enfermedad oftalmológica . . . . .	11
2.2 Piel . . . . .	11
2.2.1 Capas de la piel . . . . .	12
2.2.2 Enfermedades que afectan a la piel . . . . .	14
2.3 Tomografía de coherencia óptica . . . . .	14
2.3.1 Conceptos fundamentales . . . . .	14
2.3.2 Propiedades de la señal OCT . . . . .	17
2.3.3 Utilidad clínica . . . . .	18

## 2.1 Ojo

El ojo es el órgano base del sistema visual. Se encarga de transformar la energía lumínica en impulsos eléctricos que, tras transmitirse a la corteza visual del lóbulo occipital, permiten la percepción tridimensional del movimiento, la forma y la profundidad. En condiciones normales, el ojo presenta un peso de aproximadamente 7,5 gramos y un diámetro antero-posterior de 24 mm. Posee un ángulo de visión de 200 grados y es capaz de distinguir hasta 10 millones de colores y sombras [7].

Para comprender el funcionamiento de este órgano tan complejo y estudiar así sus posibles afecciones, es necesario conocer su anatomía, (véase Figura 2.1). A continuación, se explican las diferentes partes que componen el globo ocular, así como la función que cada una desempeña en el proceso de visión.

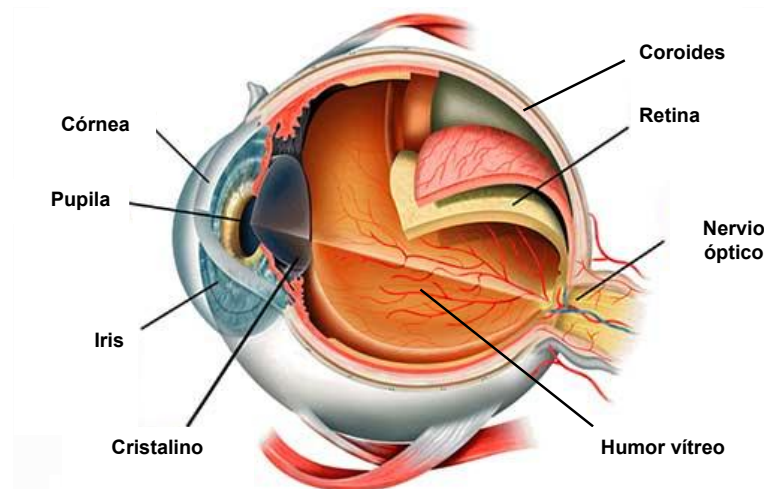


Figura 2.1: Principales componentes presentes en el globo ocular.

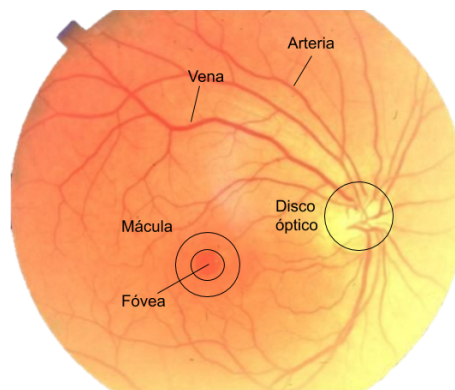
### 2.1.1 Anatomía del ojo

En la parte más profunda del interior del globo ocular, se localiza el humor vítreo, que ocupa casi dos tercios del volumen total. Se caracteriza por ser un tejido conjuntivo, transparente, esférico y recubierto por hialoides. Su principal función es mantener la forma ocular y regular el metabolismo. En la parte anterior del humor vítreo se localiza el cristalino, una lente biconvexa de +22. El cristalino se encarga de la acomodación del ojo por medio del músculo ciliar. Debido a la ausencia de su propio riego sanguíneo, el humor acuoso (líquido derivado del plasma que circula en la red capilar) se encarga de proporcionárselo [8].

Envolviendo el interior del globo ocular, se localizan tres capas fundamentales que son detalladas a continuación:

- **Capa externa.** La parte más externa de esta capa está constituida por la córnea, situada en la parte anterior del ojo y compuesta por el epitelio anterior, la membrana de Bowman, estroma, membrana Descemet y endotelio. Es transparente y presenta una curvatura regular, lo que la convierte en una lente convergente. Su principal función es la protección del polo anterior del ojo ante traumatismos. Posterior a la córnea, se encuentra la esclerótica, una capa blanca compuesta por tejido conjuntivo y fibras elásticas. Estas fibras presentan una gran resistencia, lo que convierten a esta capa en la cubierta protectora de la parte posterior del ojo [8].

- **Capa media o úvea.** El componente fundamental de esta capa es el iris, cuya función es la regulación de la cantidad de luz que penetra a través de la pupila. En esta capa, también se encuentra el cuerpo ciliar, donde se localiza el músculo ciliar y el coroides. Este último es el encargado de nutrir a la retina por medio de la multitud de vasos sanguíneos que presenta [8].
- **Capa interna o retina.** La retina es una estructura derivada del ectodermo neural especializada en recoger, elaborar y transmitir estímulos visuales. Se sitúa entre la coroides y el vítreo y, se visualiza en el fondo de ojo como una estructura transparente y rojiza, (véase Figura 2.2). En esta figura, se pueden observar las estructuras anatómicas más importantes de la retina. Entre ellas, destaca el nervio óptico. Esta estructura es la conexión entre la retina y el centro de procesamiento visual del cerebro. El área ovalada donde el nervio óptico cruza a través del fondo posterior del globo ocular se denomina disco óptico. Dado que no contiene fotorreceptores, también es llamado punto ciego. Aquí es donde la arteria central de la retina entra y la vena central sale. Dado que es el lugar de salida de las fibras ópticas, es importante conocer su forma y su situación, pues estas pueden ser indicadores de enfermedades oculares y sistémicas. Otra parte importante de la retina es la mácula. En su interior se localiza la fóvea, zona con mayor agudeza visual por su elevada concentración de conos. Por último, los vasos sanguíneos, arterias y venas presentes en la retina son los responsables de proporcionarle oxígeno y nutrientes [9].



**Figura 2.2:** Fondo de ojo donde se pueden visualizar las partes más relevantes de la retina.

El estudio en profundidad de la retina resulta fundamental ya que muchas de las patologías oftalmológicas se manifiestan en esta capa del ojo. Como se puede observar en la Figura 2.3 a nivel histológico, la retina consta de diferentes capas. Dado que uno de los objetivos del proyecto desarrollado es la segmentación de algunas de ellas para su posterior estudio, a continuación, se explican con detalle, recorriéndolas desde la más externa hasta la más interna [10].

- **Epitelio pigmentario (RPE):** está formada por una capa unicelular de células cúbicas que contienen gránulos de melanina.
- **Capa de células fotorreceptoras (OS):** está compuesta por los segmentos externos de los conos y los bastones, que contienen los discos de pigmentos visuales en su interior y forman la parte más importante de la retina. Estas células son las encargadas de la absorción de la luz y su transformación en impulsos nerviosos.
- **Capa limitante externa (IS):** compuesta por los complejos de unión entre las membranas celulares de las células de müller y la de los segmentos internos de los fotorreceptores.

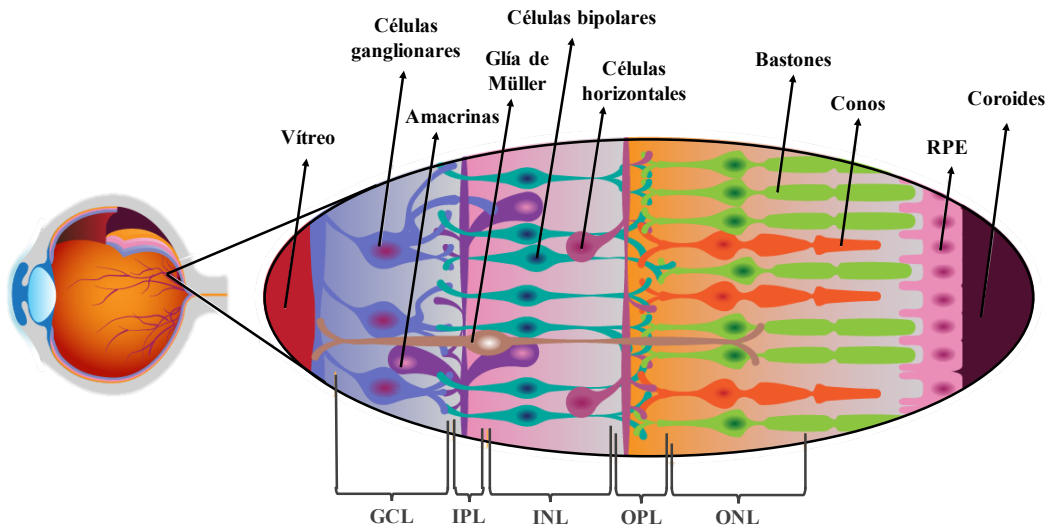


Figura 2.3: Composición de las diferentes capas que conforman la retina.

- **Capa nuclear o granular externa (ONL):** está formada por los núcleos celulares de las células fotorreceptoras, conos y bastones.
- **Capa plexiforme externa (OPL):** es la región de conexión sináptica entre células fotorreceptoras, bipolares y horizontales.
- **Capa nuclear interna (INL):** está formada por los núcleos celulares de las células bipolares, horizontales y amacrinas. Las células horizontales y amacrinas ayudan y procesan la información de los fotorreceptores.
- **Capa plexiforme interna (IPL):** es la región de conexión sináptica entre células bipolares, amacrinas y ganglionares. Además, a este nivel terminan gran cantidad de prolongaciones de las células amacrinas, que influyen y modulan la información que se transmite a las células ganglionares.
- **Capa de las células ganglionares (GCL):** está constituida por los núcleos de las células ganglionares.
- **Capa de fibras del nervio óptico (RNLF):** está formada por los axones de células ganglionares que forman el nervio óptico.
- **Membrana limitante interna (ILM):** es una membrana basal filamentosa que está formada por las extensiones de los pies de las células de Müller y que se encuentran en contacto con la hialoide posterior.

Por tanto, a nivel general, el proceso de visión comienza en las capas más externas de la retina, donde la luz es captada por el epitelio pigmentario y las células fotorreceptoras. Las moléculas foto-pigmentarias de los fotorreceptores absorben la luz con el objetivo de convertirla en impulsos eléctricos. La conversión de energía luminosa a impulsos eléctricos se lleva a cabo en las siguientes capas donde se encuentran las células ganglionares que generan un potencial de acción que comienza su camino hacia el cerebro. Este transporte se hace a través de las fibras del nervio óptico que, como se ha detallado anteriormente, está constituido por los axones de algunas de las células ganglionares. Una vez en el cerebro, específicamente en el córtex cerebral, las señales eléctricas son transformadas en percepciones visuales [9].



### 2.1.2 *Glaucoma como enfermedad oftalmológica*

El glaucoma es una neuropatía óptica progresiva causada por un aumento de la presión intraocular debido a un mal drenaje del fluido ocular. Se caracteriza por la progresiva excavación de la papila óptica, la muerte de células ganglionares pertenecientes a la capa GCL y el adelgazamiento de la capa de fibras nerviosas de la retina (RNFL) [11]. Clínicamente, estos efectos se traducen en pérdidas funcionales como el detrimento del campo visual, que progresa en una completa pérdida de la visión.

El diagnóstico del glaucoma en estadios iniciales, antes de que se produzca una pérdida funcional significativa, sigue siendo un reto y uno de los mayores obstáculos en la práctica clínica. Actualmente, el diagnóstico del glaucoma se puede realizar en base a varios estudios. La medida de la presión intraocular por tonometría, el estudio del campo visual por capimetría, la computación del radio entre la copa óptica y el área del disco óptico en las imágenes de fondo de ojo o el análisis de las imágenes pertenecientes a las capas de la retina son algunos de ellos.

La mayoría de estudios científicos sugieren que las alteraciones estructurales preceden a las pérdidas funcionales en los enfermos con glaucoma. Las alteraciones en la capa de fibras nerviosas de la retina pueden detectarse con una pérdida de un 5% de células ganglionares, mientras que la detección de la pérdida del campo visual y de la agudeza visual requieren la muerte de un 25-35% y de hasta un 40% respectivamente [12]). Es por ello que el análisis de las capas RNFL+GCL peripapilar es el protocolo de escaneo más eficaz para el diagnóstico precoz del glaucoma y, por tanto, en el que se centra este proyecto.

## 2.2 Piel

La piel es la cubierta externa del cuerpo humano y uno de los órganos más importantes debido a su tamaño y funciones. Una de sus principales funciones es la protección contra las agresiones mecánicas, químicas y de agentes patógenos. Por otra parte, también se encarga de mantener el equilibrio químico de fluidos corporales y transmitir gran cantidad de información externa. Las tres capas de tejido que componen la piel (epidermis, dermis y tejido subcutáneo) pueden ser identificadas en la Figura 2.4.

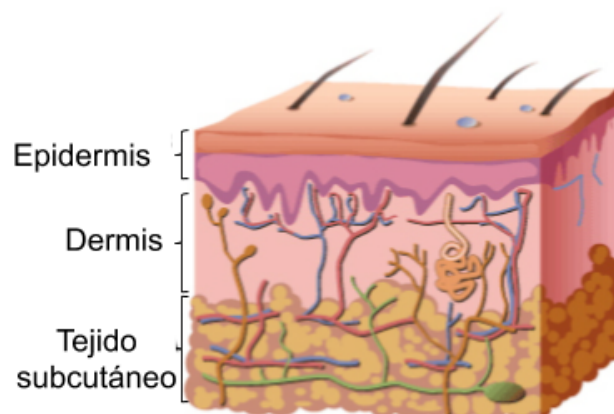


Figura 2.4: Principales capas que componen la piel.

### 2.2.1 Capas de la piel

A continuación, se describe la composición de las principales capas que forman la piel desde la más interna hasta la más externa.

#### ■ Epidermis

La epidermis es un epitelio plano poliestratificado y queratinizado con un gran número de células. A su vez, esta capa está compuesta por cinco estratos diferentes, tal y como se muestra en la Figura 2.5.

- **Capa basal:** está constituida por una capa simple de células cuboidales que normalmente se disponen en hilera. En ella se encuentran las células madre o progenitoras y melanocitos. Algunas de las células progenitoras se dividen y forman queratinocitos que suben a la superficie externa y se incorporan a las capas más superficiales. Al ascender, se alejan de los vasos de la dermis que los nutrían. Como consecuencia de esto, sus núcleos se degeneran, las células mueren y son expulsadas al exterior. Otras células madre pertenecientes a esta capa migran a la dermis y dan lugar a las glándulas sudoríparas, sebáceas y a los folículos pilosos [13].
- **Capa de células espinosas:** la constituyen células poligonales que poseen puentes intercelulares, estructuras que sirven como medio de unión entre ellas y con las capas adyacentes [14].
- **Capa granular:** los núcleos de las células de esta capa se encuentran en fase de degeneración.
- **Estrato lúcido:** está formado por 3-5 capas de células planas muertas que contienen una sustancia precursora de la queratina o la propia queratina. Este estrato solo se encuentra en algunas partes del cuerpo como los pies y las manos.
- **Estrato córneo:** está formada por células sin núcleo que se encuentran completamente rellenas de queratina. Estas células son reemplazadas por aquellas que se encuentran en estratos más profundos [13].

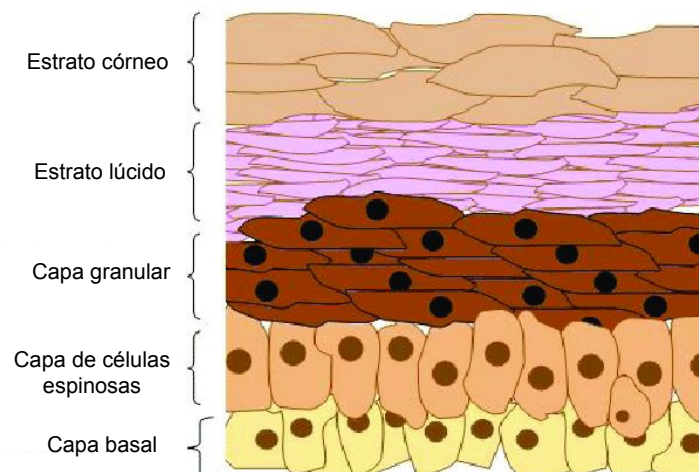


Figura 2.5: Diferentes capas pertenecientes a la epidermis.

#### ■ Dermis

Debajo de la epidermis se localiza la dermis. Esta capa está formada por la capa papilar y reticular.

- **Capa papilar:** presenta proyecciones hacia el interior de la epidermis (papilas dérmicas) donde se encuentran las asas capilares que proporcionan los nutrientes necesarios a la epidermis.
- **Capa reticular:** esta estructura proporciona elasticidad y capacidad de adaptación a movimientos y está formada por haces de fibras de colágeno entrelazados con haces de fibras elásticas.

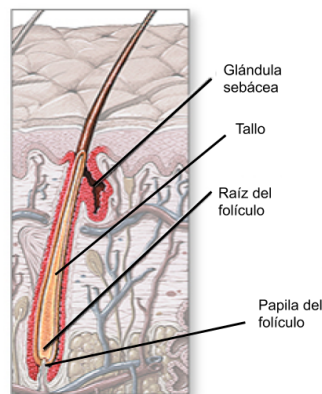
#### ■ Tejido subcutáneo

El tejido subcutáneo es una capa constituida por adipocitos que se disponen en lóbulos separados por tejido conectivo laxo. Muchas de sus fibras se fijan a las de la dermis, no existiendo un límite definido entre estas dos capas. El tejido subcutáneo proporciona el anclaje de la piel a las estructuras subyacentes como son la fascia, periostio o pericondrio. Dado que presenta tejido adiposo, esta capa sirve de almacén de energía, aislante térmico y protector mecánico frente a golpes [14].

#### ■ Anexos de la piel

A continuación, se describen algunos componentes anexos a la piel. Entre ellos destacan los pelos y las glándulas sebáceas y sudoríparas.

- **Pelos:** son filamentos queratinizados y delgados, asentados en invaginaciones epidérmicas presentes en casi toda la superficie de la piel. Tal y como se puede apreciar en la Figura 2.6, presentan dos partes principales: el tallo y la raíz o folículo piloso. El tallo está formado por tres capas concéntricas: médula del pelo, corteza o región mayor y cutícula. En la raíz, se observan células epidérmicas que forman parte de la vaina radicular interna y, otras periféricas que forman la vaina radicular externa. En su parte inferior, bulbo piloso, se encuentran las células epidérmicas que dan origen al pelo y rodean a la papila dérmica.



**Figura 2.6:** Principales partes del folículo piloso.

- **Glándulas sebáceas:** son glándulas holocrinas que producen lípidos y ayudan a mantener el manto hidrolipídico de la piel. Su conductor excretor desemboca en el folículo piloso.
- **Glándulas sudoríparas:** son glándulas tubulares que se clasifican en dos tipos diferentes: ecrinas, que controlan la temperatura y apocrinas, que producen el olor corporal característico.

### **2.2.2 Enfermedades que afectan a la piel**

El cáncer de piel se encuentra entre los tipos de cánceres más frecuentes y malignos de todo el mundo [15]. El melanoma es el tipo más agresivo de cáncer de piel, puesto que, en lo que a cáncer de piel se refiere, es el que provoca la mayoría de las muertes. Este cáncer se origina en los melanocitos, células responsables de la pigmentación de la piel, el cabello y los ojos. El melanoma se desarrolla como resultado del daño en el ADN, generalmente como consecuencia de la exposición intensa a la radiación ultravioleta. Cuando un melanocito sufre un daño genético suficiente, puede comenzar a crecer de manera anormal y convertirse en un tumor melanocítico. En la etapa más temprana del desarrollo, el tumor crece lateralmente a lo largo de la capa epidérmica de la piel. En este punto es posible la extirpación quirúrgica completa de la lesión. Sin embargo, si no se diagnostica a tiempo, el tumor puede engrosarse e infiltrarse más profundamente en la piel. Si el crecimiento del tumor invade los vasos linfáticos, las células cancerosas pueden separar la matriz del tumor primario y viajar a los ganglios linfáticos regionales como metástasis en tránsito [15]. Es por ello que el diagnóstico precoz de esta enfermedad resulta fundamental. Dado que esta patología se origina en la epidermis, es de vital importancia llevar a cabo una segmentación automática de esta capa a fin de obtener un diagnóstico temprano del melanoma.

Entre otras enfermedades importantes que afectan a la piel destaca la psoriasis. La psoriasis es una enfermedad inflamatoria crónica y multisistémica. Está caracterizada por la formación focal de placas inflamadas y elevadas que constantemente desprenden escamas derivadas del crecimiento excesivo de las células epiteliales de la piel. La enfermedad se define por una serie de cambios celulares, tales como: hiperplasia de los queratinocitos epidérmicos, hiperplasia vascular y ectasia, e infiltración de linfocitos T, neutrófilos y otros tipos de leucocitos en la piel afectada [16]. Como se ha mencionado anteriormente, este tipo de enfermedad está caracterizada por un aumento del grosor de la epidermis, por lo que en este caso, la segmentación de esta capa para la posterior medida de su grosor resulta clave a la hora de realizar un diagnóstico precoz.

## **2.3 Tomografía de coherencia óptica**

### **2.3.1 Conceptos fundamentales**

La tomografía de coherencia óptica (*Optical Coherent Tomography*) es una técnica de diagnóstico por imagen no invasiva basada en el principio de interferometría de baja coherencia. Esta modalidad utiliza la reflexión de la luz infrarroja para la obtención de cortes tomográficos. En ella, un rayo de luz es dirigido hacia el tejido y la estructura interna del mismo es medida en base al retraso del eco de luz reflejado por las microestructuras. Dado que el tipo de onda que se emplea es la luz y, esta se transmite a una velocidad muy elevada, la resolución de la OCT es alta. Este efecto permite identificar estructuras de hasta 10 micras, lo que resulta útil para medir distancias con elevada precisión [17].

El dispositivo básico en el que se basa la OCT es el Interferómetro de Michelson. Este interferómetro está compuesto por cuatro elementos básicos: una fuente de luz, un divisor de haces, un espejo de referencia y un detector como puede observarse en la Figura 2.7.

Para entender el funcionamiento del interferómetro, en primer lugar, se exponen algunas características de la luz que resultan fundamentales.

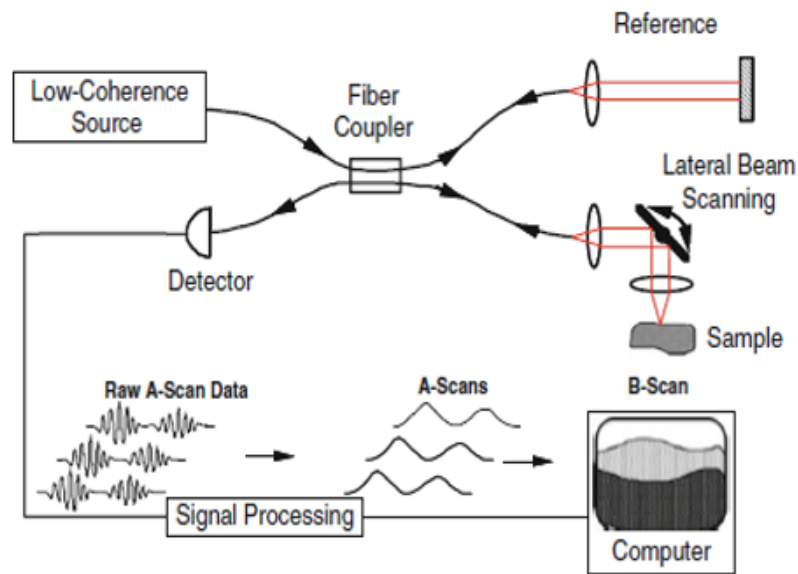


Figura 2.7: Funcionamiento del Interferómetro de Michelson [18].

Un haz de luz está compuesto por un campo eléctrico y uno magnético que oscila periódicamente en el tiempo. Cuando se combinan dos rayos de luz, los campos magnéticos y eléctricos que los componen pueden combinarse de forma constructiva o destructiva en función de la fase relativa de sus oscilaciones. Este fenómeno se conoce con el nombre de interferencia. Si se produce una interferencia constructiva y coinciden las fases de los rayos, el rayo resultante es más intenso. Sin embargo, si se produce una interferencia destructiva, el rayo resultante será menos intenso, e incluso puede alcanzar el valor 0 [19]. El fenómeno de interferencia descrito, Figura 2.8, es fundamental para la obtención de imágenes en OCT.

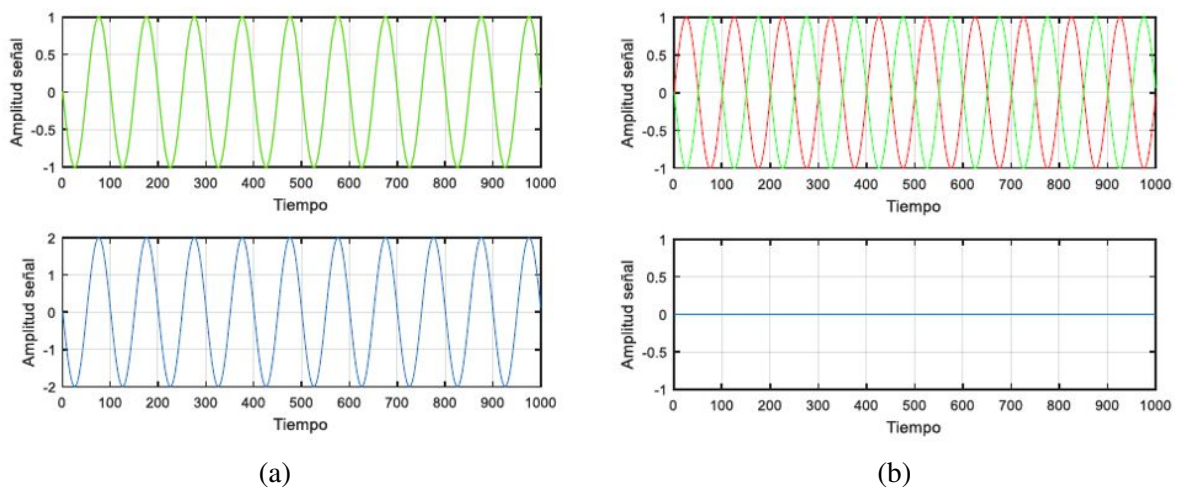
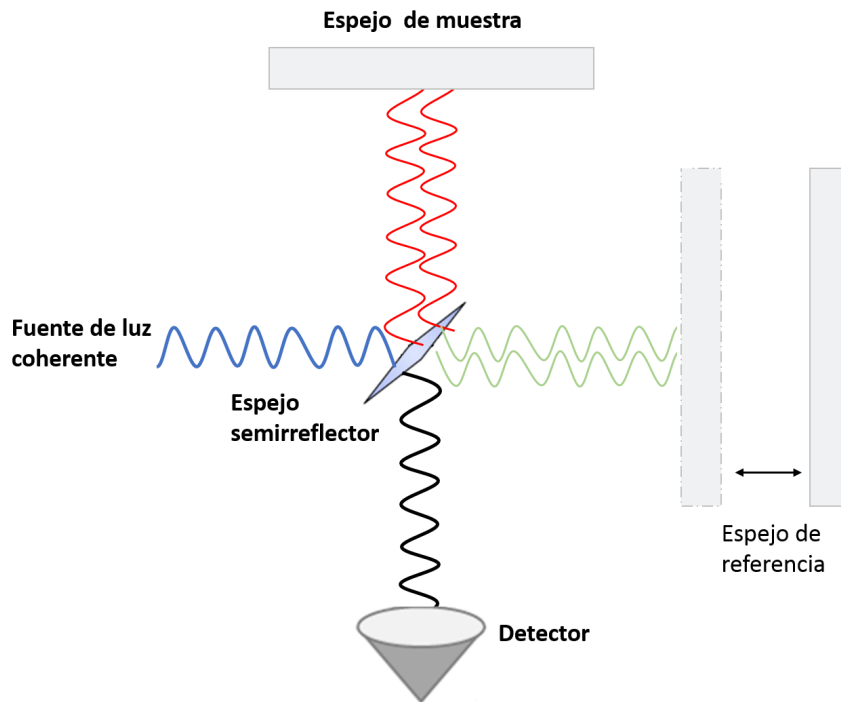


Figura 2.8: Tipos de interferencias cuando se combinan dos haces de luz. a) Interferencia constructiva: se suman las amplitudes de los dos haces de luz. b) Interferencia destructiva: se restan las amplitudes de los dos haces.

En una OCT, un haz luminoso con una banda ancha en el espectro infrarrojo de 820 a 830 nm (originado por la fuente de luz) es dirigido a un espejo parcialmente reflectante que lo divide en dos rayos. Uno de esos rayos, rayo señal (onda roja), se dirige al tejido del cual se quiere obtener la imagen y el otro,

rayo de referencia (onda verde), se refleja hacia el espejo móvil de referencia cuya posición es conocida, (véase Figura 2.9).



**Figura 2.9:** Recorrido de los haces de luz en una OCT. Haz de entrada (azul), Rayo señal (rojo), Rayo referencia (verde), Haz recombinado (negro).

El rayo señal, reflejado por la muestra hacia el espejo semirreflector o divisor de haz contiene múltiples “ecos” que son producto de las diferentes interfases con las que ha interactuado dentro del tejido. Por otro lado, el haz de referencia, verde, se refleja de vuelta también hacia el divisor de haz [20]. Sin embargo, este haz no ha sufrido modificación alguna, a excepción de un retardo variable por la modificación de la posición del espejo de referencia. Este retardo puede ser calculado en base a la velocidad del haz de luz en el medio en el que se propague y el desplazamiento del mismo [21].

Estos dos haces reflejados se recombinan en el divisor del haz. Cuando ambos coinciden en el tiempo, se produce el fenómeno de interferencia citado anteriormente. Posteriormente, el rayo resultante llega al detector y, a través de su potencia y retardo, se puede conocer el índice de refracción y la posición de cada estructura presente en el tejido de estudio [22].

Para adquirir una imagen tomográfica de una muestra, es necesario realizar dos tipos de barrido, uno axial y otro lateral. El espejo en el cual es reflejado el haz de referencia realiza un movimiento lineal que ajusta el retardo de la rama de referencia. De esta forma, se pueden medir las variaciones que se producen en la radiación cuando un haz de luz penetra en una parte del tejido, lo que se denomina registro axial (A-scan). Por otra parte, es necesario realizar un barrido lateral sobre el tejido para que se registren múltiples señales en sentido axial (A-scan) y obtener así un corte transversal del mismo (B-scan). Realizando varios B-scan en paralelo se genera un cubo en tres dimensiones que permite obtener el volumen del tejido a analizar (C-scan), (véase Figura 2.10).

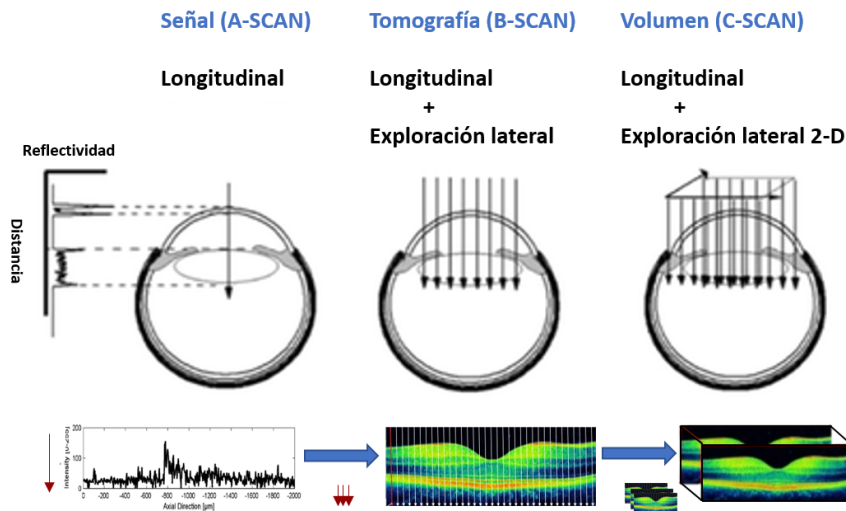


Figura 2.10: Diferencia ente A-scan, B-scan y C-scan

### 2.3.2 Propiedades de la señal OCT

En la OCT la resolución axial y transversal están determinadas por diferentes mecanismos físicos. Mientras que las medidas axiales son producto de las propiedades del tejido a lo largo del haz, las medidas transversales subyacen de la adquisición de diferentes medidas axiales en el objeto a estudiar [22].

La resolución axial,  $\{\delta L\}$ , en la dirección de propagación del haz, está determinada por la longitud de coherencia de la fuente de luz [21]. Para esta aplicación, como se trata de una luz de baja coherencia, únicamente puede ser caracterizada por su frecuencia o ancho de banda de longitudes de onda ( $\delta\lambda$ ) [17]. La relación entre la resolución axial y el ancho de banda de longitudes de onda está dada matemáticamente por la teoría de la termodinámica, ecuación 2.1.

$$\delta L = \frac{2\ln(2)}{\pi} \frac{\lambda^2}{\delta\lambda} \quad (2.1)$$

siendo  $\lambda$  la longitud de onda principal de la fuente de luz.

Por ejemplo, la resolución axial ideal de un haz luminoso con una longitud de onda  $\lambda = 800 \text{ nm}$  y un ancho de banda  $\delta\lambda = 20 \text{ nm}$  será aproximadamente de  $15 \mu\text{m}$ . Sin embargo, dado que el índice de refracción del tejido a analizar será mayor que el existente en el aire, la velocidad de la luz es menor, reduciéndose la resolución hasta aproximadamente  $10 \mu\text{m}$ . Se puede observar que se requieren fuentes de luz con un ancho de banda amplio para conseguir que la resolución axial sea elevada, ya que esta es inversamente proporcional al ancho de banda de la fuente de luz.

La resolución transversal,  $\{\delta x\}$ , mantiene una dirección perpendicular al haz y depende del tamaño del punto focalizado [17]. Se expresa con la siguiente fórmula:

$$\delta x = \frac{4\lambda}{\pi} \left\{ \frac{f}{d} \right\} \quad (2.2)$$

donde  $d$  es  $\frac{1}{e^2}$  el diámetro del haz gaussiano en la lente de enfoque y  $f$  es la longitud focal de la misma lente. Se puede conseguir una gran resolución transversal empleando una gran apertura numérica para disminuir el tamaño del punto en el plano focal de la lente [21].

### 2.3.3 Utilidad clínica

La OCT es particularmente atractiva en aplicaciones oftalmológicas dada la facilidad con la que la luz alcanza las estructuras oculares. Además, al ser una prueba no invasiva y que por tanto, no requiere el contacto directo con el ojo, permite estudiar in vivo las estructuras oculares [17]. Menos establecida está la OCT en el campo de la dermatología, donde la dispersión de la luz limita fuertemente la penetración en la piel. Sin embargo, los adelantos producidos en esta modalidad han impulsado estudios in vivo que brindan información sobre una variedad de enfermedades de la piel [23].

A continuación, se realiza una descripción de la imágenes de OCT tanto de ojo como de piel manifestando su utilidad para la detección de enfermedades.

#### ■ *Importancia de la OCT en la detección de enfermedades oculares*

La OCT permite diferenciar estructuras histológicas retinianas y subretinianas usando un haz de luz. Cuando el haz de luz llega a las capas tisulares retinianas sufre diferentes fenómenos entre los que se encuentran la absorción, transmisión y dispersión. La luz dispersada en dirección opuesta a la emitida se denomina luz reflejada. Como se ha explicado anteriormente, este último fenómeno proporciona información sobre las estructuras presentes en el objeto de estudio. En el caso de la retina, cada capa tiene diferentes propiedades que definen su mayor o menor reflectividad. La OCT permite medirla y representarla mediante una escala de grises o de colores [24].

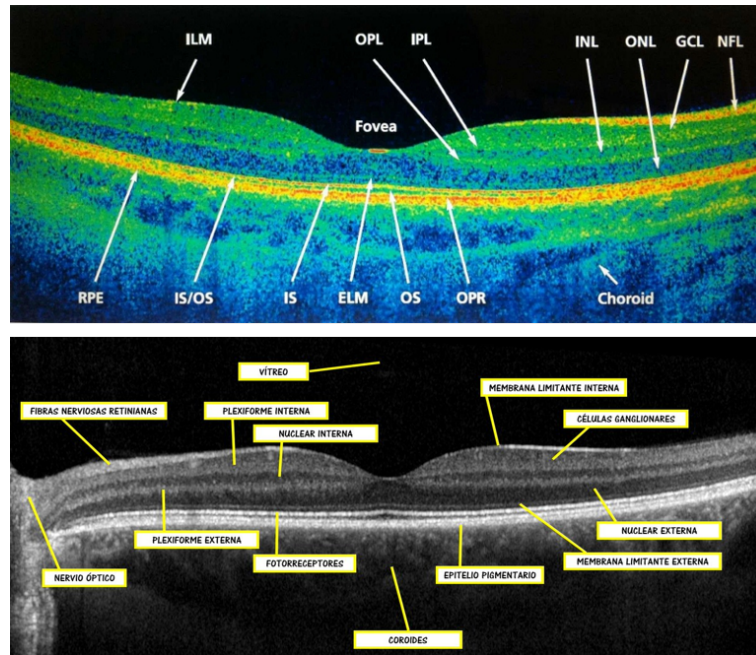
La elevada reflectividad se suele representar mediante los colores rojo o blanco, en caso que sea escala de grises, e indica que el tejido tiene una gran reflexión, por lo que bloquea la transmisión de la luz. Esta propiedad es característica de zonas que presentan fibrosis, sangre y exudados lipídicos, del epitelio pigmentario de la retina, de la coriocalpilar, etcétera.

La banda de reflectividad media, verde y amarillo en la Figura 2.11, corresponde a las capas que van desde la membrana limitante interna (INL) a la plexiforme externa (OPL).

La baja reflectividad se suele representar mediante los colores azul o negro, en caso de escala de grises. Este fenómeno indica que el tejido posee muy poca reflexión de la luz y es característico de las zonas de edema, de las cavidades quísticas de contenido seroso, del vítreo, y de la capa de fotorreceptores. En la Figura 2.11 se pueden observar las características anteriormente expuestas.

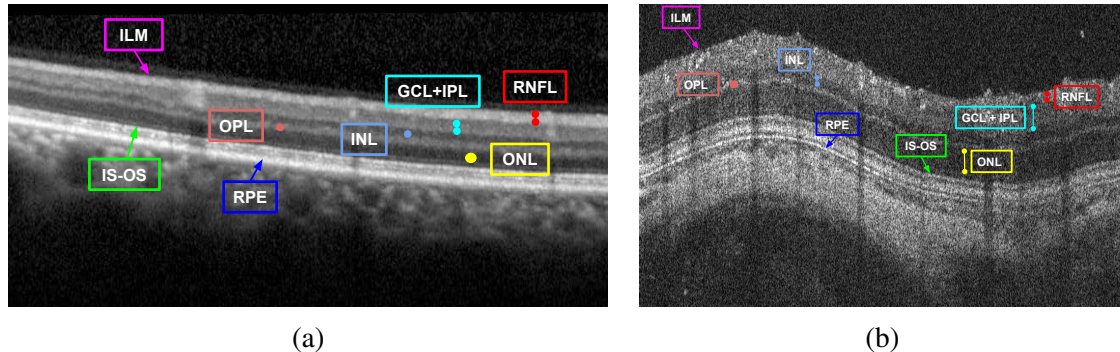
Dado que la OCT proporciona imágenes de la retina de gran resolución, los médicos las utilizan para detectar trastornos de la retina, incluyendo degeneración macular, glaucoma y trastornos que pueden causar el desarrollo de nuevos vasos sanguíneos en el ojo. Además, la OCT se puede adaptar para la adquisición de imágenes de roedores con el objetivo de utilizarlas en estudios de investigación de enfermedades oftalmológicas, así como en la evaluación de diferentes tratamientos. Sin embargo, se debe tener en cuenta que la retina humana y la de roedores presentan diferencias significativas. Por ejemplo, los roedores no tienen mácula ni fovea y sus lentes son relativamente más grandes. Con respecto al grosor de las capas retinianas, la capa de células ganglionares (GCL) en roedores no es distinguible visualmente (aproximadamente  $2\mu\text{m}$ ), mientras que en humanos toma valores alrededor de  $20\text{-}60\mu\text{m}$ . La tercera capa más interna, la capa plexiforme interna (IPL), por el contrario, es relativamente gruesa





**Figura 2.11:** Imágenes OCT a color y escala de grises de la retina donde se pueden observar las diferentes capas que la componen.

en la retina de la rata (aproximadamente  $60 \mu\text{m}$ ). En la Figura 2.12, se pueden observar las diferencias entre las imágenes de OCT de humanos y roedores.



**Figura 2.12:** Capas de la retina en dos scanners: (a) Retina humana, Spectralis OCT (Heidelberg Eng.). (b) Retina de un roedor, Micron IV (Phoenix Research Labs)

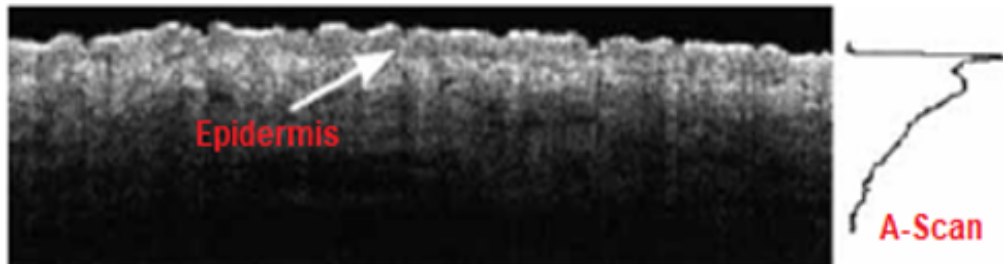
Como se explicará en secciones posteriores, el estudio realizado en este proyecto se ha centrado en las imágenes de retina que pertenecen a roedores.

#### ■ **Importancia de la OCT en la detección de enfermedades epiteliales**

El diagnóstico y tratamiento de la mayoría de las enfermedades de la piel es efectuado mediante la extracción de una biopsia y su posterior análisis histológico. Sin embargo, estas técnicas son invasivas y dolorosas. Para lidiar con estas desventajas, en los últimos años la aplicación de la OCT en piel ha adquirido una mayor relevancia. Esta modalidad de diagnóstico permite obtener imágenes de piel de hasta 2 mm de profundidad, aumentándola si la piel es tratada con glicerol. Esta profundidad es suficiente para analizar muchas de las enfermedades que afectan a la piel [25].

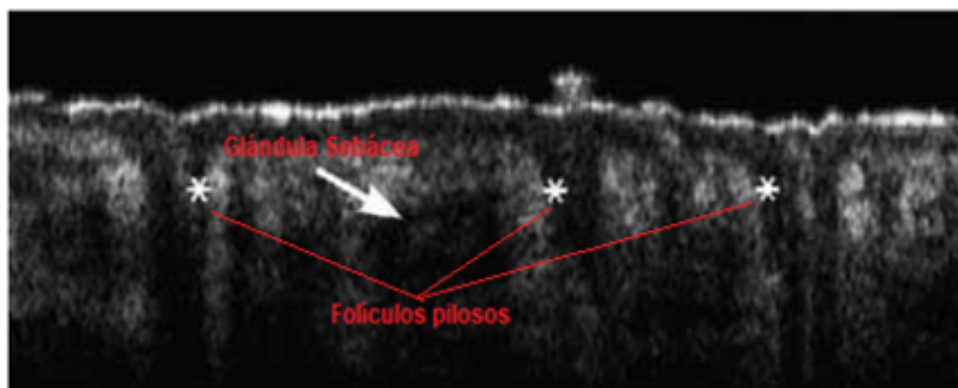
Al visualizar la piel mediante OCT, su elevada reflectividad conduce a un pico de intensidad de señal en la superficie. Por debajo de esta señal de entrada, se pueden observar las estructuras que la OCT es capaz de distinguir.

Como se ha explicado en la sección 2.2, la epidermis es la primera capa de la piel. En una imagen de OCT, su intensidad es menor que la de la dermis. En el A-scan promediado, el borde de la dermis se caracteriza por un segundo pico de intensidad posterior al de la epidermis. En la Figura 2.13 se encuentra ejemplificado este fenómeno.



**Figura 2.13:** Epidermis visualizada en una imagen OCT de piel

La dermis muestra señales intensas con algunas regiones reflectantes más bajas, correspondientes a los folículos pilosos y las glándulas sebáceas. Los vasos sanguíneos aparecen como estructuras redondas o alargadas [26]. Si la dermis es delgada, el tejido subcutáneo de baja reflexión es visible en las partes inferiores de la imagen. Estas características se pueden apreciar en la Figura 2.14.

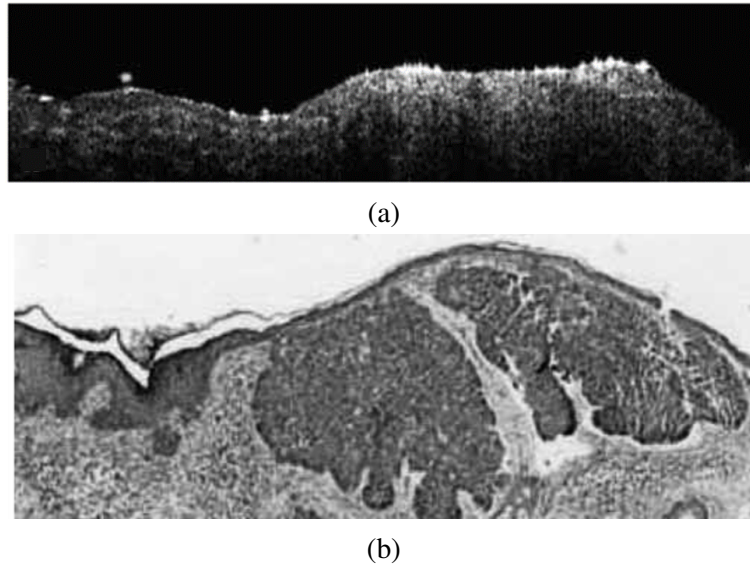


**Figura 2.14:** Folículos pilosos y glándula sebácea visualizados en una imagen OCT de piel.

A continuación, se explican los cambios morfológicos que se producen en la piel como consecuencia de padecer determinadas enfermedades y cómo estos pueden distinguirse en una imagen de OCT.

En imágenes OCT de tumores epiteliales, la derivación de los agregados de células tumorales de la epidermis es visible. En estas imágenes los tumores sólidos muestran una distribución homogénea en la señal. Este fenómeno puede ser observado en la Figura 2.15 (a).

Por otro lado, las estructuras quísticas pueden ser identificadas por áreas libres de señales. Los tumores melanocíticos de la piel muestran una mayor dispersión de la luz y, al igual que otros tumores, una distribución de la señal más homogénea en comparación con la piel sana. El segundo pico de intensidad, que representa una frontera intacta entre la epidermis y la dermis, desaparece en el melanoma maligno infiltrante en crecimiento [26].



**Figura 2.15:** Imágenes tumorales de piel. a) Imagen OCT, b) Imagen histológica correspondiente.

Con respecto a enfermedades inflamatorias de la piel, en una OCT se pueden observar los cambios que estas producen en la morfología de la piel. El estrato córneo paraqueratótico aparece como una señal de entrada engrosada de varias capas. Segmentando la capa epidérmica, la hiperplasia epidérmica difusa (engrosamiento de la piel) debida a la proliferación o espongiosis puede ser observada y cuantificada a lo largo del tiempo [26].

En la piel psoriásica, el segundo pico de intensidad es menor que en la piel sana adyacente. En la piel eczematosa, la superficie muestra más pliegues. Al igual que en la psoriasis, la señal de entrada es pronunciada y más gruesa debido a la paraqueratosis.

Por otra parte, esta modalidad permite hacer el seguimiento del tratamiento tópico<sup>1</sup> en las enfermedades anteriormente citadas, ya que sus consecuencias influyen en el aspecto óptico de la piel en la OCT. Después de algunos minutos, se puede observar la penetración del producto en el estrato córneo. Los efectos hidratantes de las cremas sobre el estrato córneo, así como la atrofia epidérmica después del tratamiento con esteroides, pueden cuantificarse mediante mediciones del espesor. Por tanto, la OCT permite la monitorización no invasiva de la curación de las enfermedades inflamatorias de la piel bajo tratamiento.

En conclusión, en el estudio de las enfermedades de la piel, el grosor de la epidermis es un parámetro muy importante. El espesor anormal de esta capa implica enfermedades tales como la psoriasis, eczema y cambios displásicos precancerosos. Es por ello que, uno de los principales objetivos de este proyecto es la correcta obtención de esta capa, haciendo uso de métodos de segmentación automática basados en algoritmos de *deep learning*.

---

<sup>1</sup>El tratamiento tópico es aquel que permite que el principio activo se aplique directamente sobre el órgano diana, en este caso la piel, con el fin de normalizar la reproducción celular excesiva y reducir la inflamación (el enrojecimiento).



## Capítulo 3

# Materiales y métodos

### *Índice de contenidos*

3.1 Materiales . . . . .	24
3.1.1 Bases de datos. . . . .	24
3.1.2 Software utilizado. . . . .	27
3.1.3 Hardware . . . . .	27
3.2 Metodología . . . . .	28
3.2.1 Estado del arte de las redes neuronales convolucionales . . . . .	29
3.2.2 Principales capas que componen una red neuronal convolucional . . . . .	31
3.2.3 Entrenamiento de los algoritmos de segmentación de capas retinianas y epiteliales. . . . .	36
3.2.4 Segmentación de las capas de la retina y piel mediante encoder-decoder . . . . .	39
3.2.5 Generación del ground truth . . . . .	42
3.2.6 Partición de los datos . . . . .	47
3.2.7 Testeo de los algoritmos de segmentación de capas retinianas y epiteliales . . . . .	48
3.2.8 Post-procesado . . . . .	49

## 3.1 Materiales

### 3.1.1 Bases de datos

Para el desarrollo de este proyecto se ha contado con dos bases de datos distintas, una perteneciente a la retina de diversos roedores, y otra, a la piel de distintos sujetos humanos. A continuación se describe cada una de ellas con profundidad.

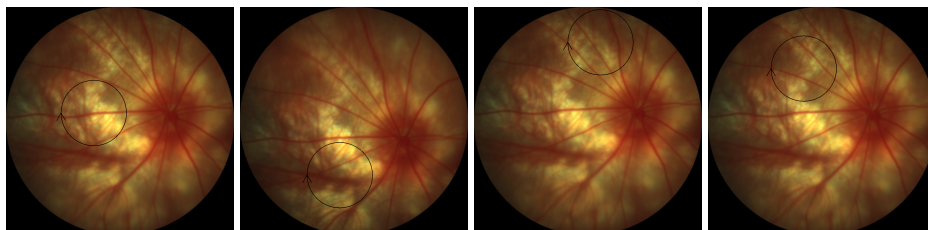
#### ■ Base de datos retiniana

Para abordar uno de los objetivos de este proyecto y generar un algoritmo automático de segmentación de las capas retinianas, se han utilizado dos entregas de una base de datos de imágenes de OCT procedente de roedores (ratas). La primera entrega se ha utilizado con propósitos de entrenamiento y validación del algoritmo propuesto, mientras que la segunda ha sido destinada a establecer un segundo testeo del algoritmo. Con ello, se pretende comprobar si el modelo propuesto, tras haber sido entrenado con unas determinadas imágenes, es capaz de obtener una segmentación precisa de nuevas imágenes correspondientes al mismo tipo de sujeto, en este caso roedores.

Todas las imágenes pertenecientes a las diferentes entregas de la base de datos fueron adquiridas con el equipo Micron IV (Phoenix Research Labs, Pleasanton, EE. UU) en diferentes instantes temporales. En ambos casos, el Consejo Danés para la Experimentación con animales otorgó el permiso de experimentación con animales con el número de licencia: 2017-15-0201-0121.

Las imágenes se adquirieron antes y después de una inyección intravítrea de endotelina-1 (ET-1) a los roedores. ET-1 es un vasoconstrictor de los vasos retinianos que causa isquemia. Por lo tanto, esta inyección contribuye a la degeneración de las capas retinianas, simulando condiciones de enfermedades de la retina como el glaucoma. Este hecho hace que el modelo de inyección intravítrea ET-1 en roedores sea adecuado para estudios de investigación sobre la monitorización de la capa retiniana.

En particular, para la primera entrega de la base de datos, se realizó un seguimiento 3, 7 y 14 días después de la inyección de ET-1. Se adquirieron un total de 129 imágenes OCT de 1024x1024 píxeles con una resolución espacial de 0.9775  $\mu\text{m}/\text{píxel}$ . Estas imágenes pertenecen a un conjunto de 10 ratas Sprague-Dawley. Con respecto a las imágenes de la segunda entrega de la base de datos, se realizó un seguimiento 2, 3 y 7 días después de la inyección, con lo que se obtuvo un total de 115 imágenes OCT con la misma resolución que las de la primera entrega. Estas imágenes se adquirieron a partir de 12 roedores diferentes. En promedio, se registraron 13 imágenes por roedor para la primera entrega y 10 para la segunda de exploraciones circulares ubicadas en diferentes posiciones del fondo del ojo tal y como puede ser observado en la Figura 3.1.



**Figura 3.1:** Ejemplificación de las distintas posiciones en las que se han adquirido las imágenes de la base de datos de roedores.

Estas exploraciones circulares consisten en la generación de una imagen (B-scan) mediante la adquisición de múltiples (A-scan) alrededor de los círculos observados en la Figura 3.1.

A modo resumen, en las tablas 3.1 y 3.2 se detalla el contenido de las dos entregas de la base de datos, indicando el número de imágenes que pertenecen a cada roedor, así como las que fueron adquiridas en cada momento.

ID del Roedor	Nº de imágenes	Día 0	Día 3	Día 7	Día 14
C23 GP1A A1	16	4	4	4	4
C23 GP1A A2	17	4	5	4	4
C23 GP1A A3	16	4	4	4	4
C23 GP1A A4	16	4	4	4	4
C24 GP2B A1	16	4	4	4	4
C24 GP2B A2	16	4	4	4	4
C24 GP2B A3	16	4	4	4	4
C24 GP2B A4	4	4	0	0	0
C25 GP1A A1	8	4	4	0	0
C25 GP2B A3	4	4	0	0	0

**Tabla 3.1:** Contenido de la primera entrega de la base de datos de roedores. Antes de la inyección: Día 0. Después de la ET-1 inyección: Día 3, Día 7 y Día 14.

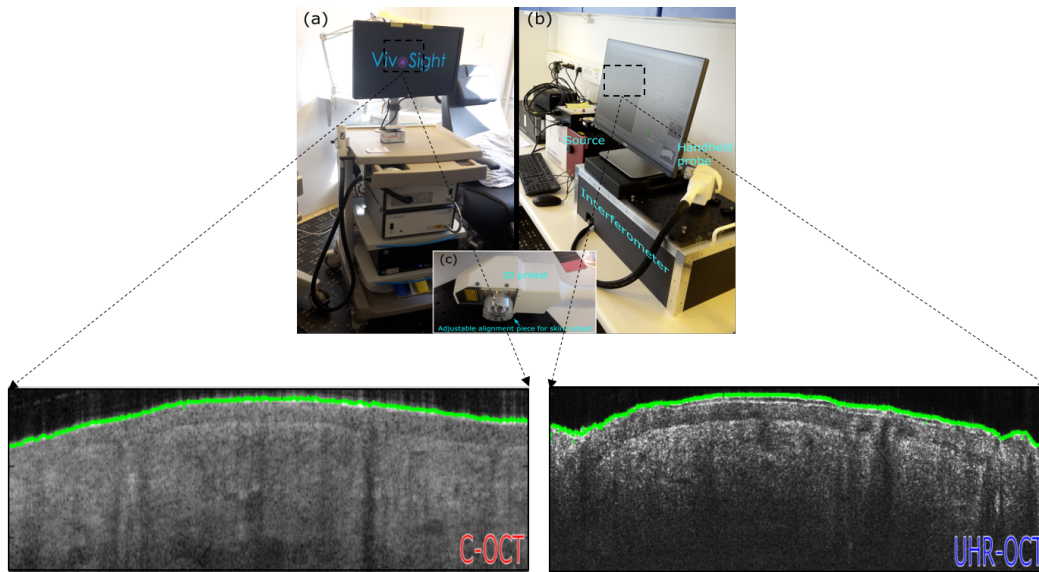
ID del Roedor	Nº de imágenes	Día 0	Día 3	Día 7
Rat01	7	5	2	0
Rat02	11	5	3	3
Rat03	10	5	3	2
Rat04	7	5	0	2
Rat05	11	5	2	4
Rat06	10	5	2	3
Rat07	11	4	4	3
Rat08	9	4	4	4
Rat09	14	5	4	5
Rat10	7	4	2	1
Rat11	10	3	4	3
Rat12	8	4	2	2

**Tabla 3.2:** Contenido de la segunda entrega de la base de datos de roedores. Antes de la inyección: Día 0. Después de la ET-1 inyección: Día 3 y Día 7.

### ■ Base de datos de piel

Para llevar a cabo otro de los objetivos de este proyecto y generar un modelo para segmentar la capa epitelial de la piel en imágenes de OCT, se ha utilizado una base de datos compuesta por imágenes de piel. Esta base de datos cuenta con un total de 270 imágenes pertenecientes a 9 sujetos diferentes. En este caso, todos los sujetos son humanos, sin ningún tipo de enfermedad y con edades comprendidas entre 25 y 75 años. El análisis de imágenes de pacientes sanos puede ser muy importante para mejorar los caracteres de imágenes no patológicas y establecer las tendencias y puntos de referencia [23].

Las imágenes fueron adquiridas de la mejilla de los sujetos por un sistema UHR-OCT, (ultrahigh resolution), desarrollado en Dinamarca por colaboradores del proyecto GALAHAD. En la Figura 3.2, se comparan las imágenes producidas por un sistema de OCT comercial (Vivosight de Michaelson Diagnostics) y por el sistema OCT desarrollado. Como se puede observar, la mejora de las resoluciones axial y lateral que introduce este sistema, se refleja en la mejor apreciación de los detalles de la imagen de piel en OCT.



**Figura 3.2:** Imágenes de OCT de la mejilla donde las delimitaciones verdes marcan el trazado de la superficie realizado. a) sistema OCT comercial b) sistema UHR-OCT c) sonda de mano del sistema UHR-OCT.

En concreto, la resolución espacial de las imágenes adquiridas con el sistema UHR-OCT, es decir, la distancia física entre los píxeles, es aproximadamente de  $2 \mu\text{m}$  en cada dirección. Con respecto a la resolución óptica, esta tiene un valor de  $4 \mu\text{m}$  en la dimensión axial (dentro de la piel) y  $5\text{-}6 \mu\text{m}$  en la dimensión lateral.

En cuanto al tamaño de las imágenes, en cada paciente el número de píxeles es distinto, tal y como se puede apreciar en la Tabla 3.3.

Pacientes	Nº de imágenes	Dimensiones
Paciente01	30	975 x 601
Paciente02	30	975 x 601
Paciente03	30	975 x 601
Paciente04	30	975 x 526
Paciente05	30	975 x 601
Paciente06	30	995 x 601
Paciente07	30	995 x 601
Paciente08	30	995 x 551
Paciente09	30	995 x 601

**Tabla 3.3:** Número y dimensiones de las imágenes que pertenecen a la base de datos de piel.

Dado que el objetivo de la utilización de estas imágenes es el desarrollo de un algoritmo de segmentación basado en *deep learning* y, para la implementación de este tipo de algoritmos todas las imágenes deben tener las mismas dimensiones, se ha hecho un reescalado de todas las imágenes a la dimensión mínima, en este caso  $975 \times 526$ .



### **3.1.2 Software utilizado**

Para la implementación de los algoritmos propuestos y el análisis de los resultados obtenidos en este proyecto, se ha empleado el entorno de programación MATLAB® v.R2018a. El software MATLAB®, cuyo nombre proviene de *MATrix LABoratory*, es una plataforma desarrollada para resolver problemas relacionados con la rama de la ingeniería y las ciencias que incorpora un lenguaje de alto nivel basado en operaciones con matrices. Se utiliza para el desarrollo de aprendizaje automático y visión artificial, procesamiento de señales e imágenes, finanzas computacionales y otros muchos campos. Además, se encuentra disponible para las plataformas Unix, Windows, Mac OS X y GNU/Linux [27].

Para la implementación de las arquitecturas basadas en redes neuronales convolucionales, se hecho uso de la librería MatConvNet de Matlab v.1.0-beta25. Esta herramienta está basada en la simplicidad y la flexibilidad, exponiendo los componentes básicos de las CNNs como funciones de Matlab sencillas de implementar [28].

Por último, para la etapa de entrenamiento de los diferentes algoritmos de segmentación basados en *deep learning*, se hace uso de la librería NVIDIA Cuda® Deep Neural Network (cuDNN) [29]. Dado que este entrenamiento se ha ejecutado en un servidor remoto por las altas prestaciones computacionales que se requieren, se ha hecho uso del programa MobaXterm [30]. Esta herramienta permite generar una conexión SSH (Secure SHell) facilitando la comunicación segura entre el cliente (el ordenador de trabajo) y el servidor (donde se ejecutan los distintos algoritmos).

### **3.1.3 Hardware**

La mayor parte del proyecto ha sido desarrollado en un equipo compuesto por un procesador intel® Core™ i7-4770K @3.4 GHz con un sistema operativo Windows 10 de 64 bit. En cuanto a la memoria RAM, la capacidad es de 16 GB y, además, el equipo cuenta con una tarjeta gráfica NVIDIA GeForce GTX650.

Como se ha expuesto anteriormente, para el entrenamiento de las CNNs, se requieren servidores que cuenten con procesadores más potentes y con una mayor capacidad de almacenamiento. Es por ello que, para llevar a cabo este proceso, se ha hecho uso de un servidor de computación de altas prestaciones perteneciente al grupo de investigación CVB Lab. Este está compuesto por un procesador Intel i7 @4.20GHz, 32GB de RAM y una tarjeta gráfica NVIDIA Titan V.

Por otro lado, dado que en este proyecto se ha trabajado con gran cantidad de imágenes que ocupan espacio en disco, se ha hecho uso del servidor Synology DS416 de 16 TB de capacidad para almacenarlas. Así se ha establecido una diferencia entre el servidor de computación, donde se han almacenado y ejecutado los scripts correspondientes al código implementado y, el servidor de datos, en el que se han almacenado todos los datos necesarios para llevar a cabo el proceso de entrenamiento, validación y testeo de los algoritmos.

### 3.2 Metodología

Con el objetivo de presentar de forma clara las diferentes fases que componen esta sección, en la Figura 3.3 se expone un diagrama de flujo donde se pueden observar las diferentes etapas correspondientes al proceso de segmentación para la detección de las estructuras objetivo. Cabe destacar que dicho diagrama de flujo ejemplifica los pasos llevados a cabo tanto para la segmentación de imágenes de piel como de retina con variantes en la implementación de cada paso. Todas estas etapas se explicarán de forma detallada durante el desarrollo de dicha sección.

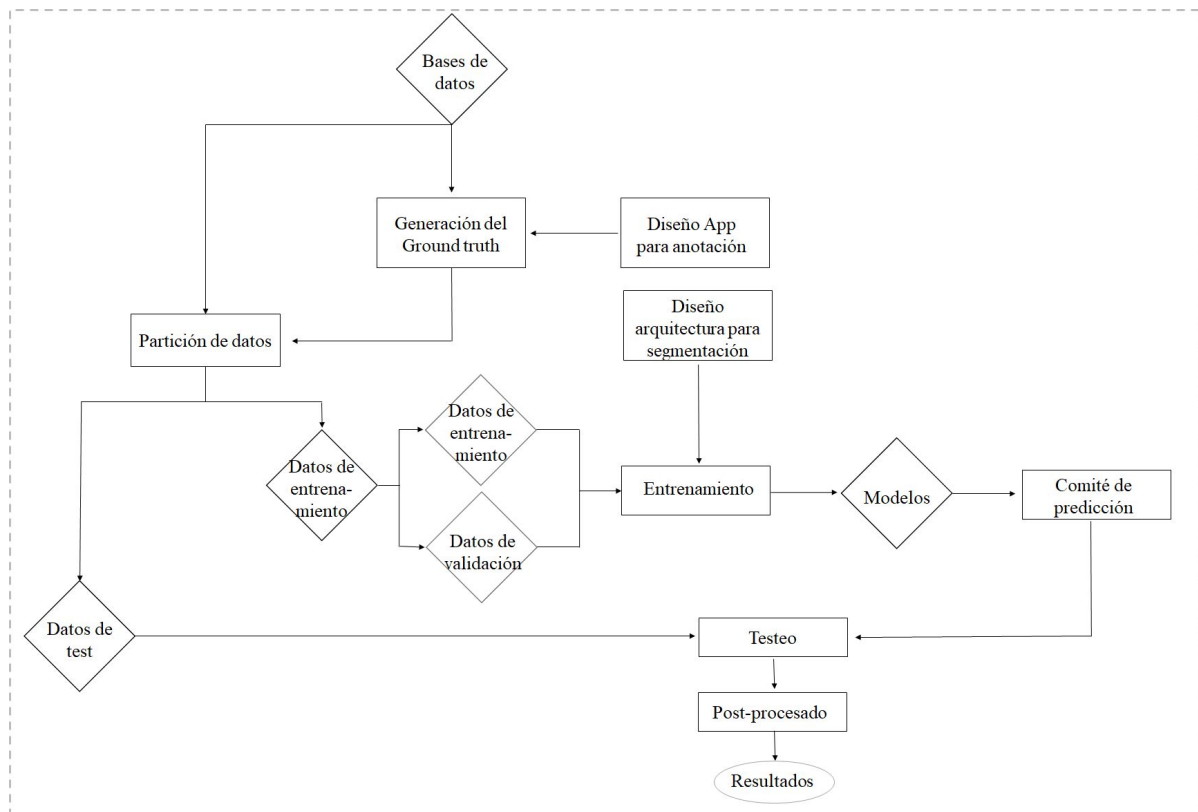
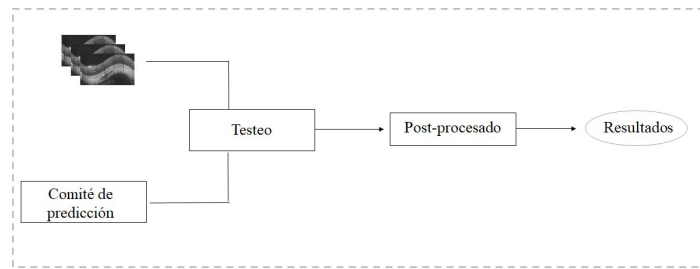


Figura 3.3: Diagrama de flujo que refleja los procesos realizados en este proyecto.

Una vez detallados los procesos desarrollados para la construcción de los modelos y su testeo, se lleva a cabo una segunda validación del método propuesto para la segmentación de imágenes de retina con nuevos sujetos<sup>1</sup>. Este hecho ejemplifica la implementación de dichos modelos en la práctica clínica real. En este caso, el diagrama de flujo se puede observar en la Figura 3.4.

<sup>1</sup>Solo se realizó un segundo testeo con imágenes de retina ya que no se ha contado con otra entrega de la base de datos de piel.



**Figura 3.4:** Diagrama de flujo que refleja el segundo proceso de validación efectuado sobre la segunda entrega de la base de datos retiniana.

### 3.2.1 Estado del arte de las redes neuronales convolucionales

Una red neuronal convolucional es un tipo de red neuronal artificial (RNA) que emplea capas convolucionales para extraer automáticamente a partir de los datos de entrada (imágenes en este caso) características discriminantes capaces de resolver ciertos problemas de clasificación o segmentación [31].

La primera arquitectura de red neuronal convolucional implementada fue denominada “LeNet”. Esta se diseñó para tareas de reconocimiento de patrones. Posteriormente, la arquitectura diseñada por Krizhevsky et al., AlexNet, fue el primer algoritmo de red neuronal profunda en ganar el Desafío de Reconocimiento Visual a Gran Escala de ImageNet (ILSVRC) [9]. Esta arquitectura se basó en la repetición de múltiples operaciones de convolución y diezmo con el fin de extraer características cada vez más complejas. Desde entonces, estas competiciones de análisis de imágenes se ganan de manera consistente mediante algoritmos basados en redes neuronales profundas. Sin embargo, entrenar una CNN de estas características from scratch (desde cero) no es tarea fácil. Dado que una CNN es un tipo de algoritmo de aprendizaje supervisado, requiere gran cantidad de datos etiquetados para su entrenamiento. La anotación de imágenes es una tarea tediosa y requiere gran cantidad de tiempo. Además, el entrenamiento de estos algoritmos requiere el uso de sistemas computacionales de alto rendimiento, tales como GPUs o clústers distribuidos a gran escala. Estas premisas se convierten en el cuello de botella para la utilización de estos algoritmos en muchas aplicaciones médicas.

En los últimos años, el aprendizaje profundo ha tenido un impacto sustancial en el campo de la investigación de imágenes biomédicas. El uso de las CNNs ha ganado gran popularidad en varias modalidades, como la radiografía, imágenes de resonancia magnética, ultrasonido y tomografía computarizada [32]. Específicamente, para la tarea de segmentación en imágenes de OCT, algunos estudios recientes han propuesto el uso de redes neuronales convolucionales, ya que suponen una potente herramienta en los campos de visión artificial a la hora de llevar a cabo tareas de reconocimiento de patrones [9].

Con respecto a la segmentación de las diferentes capas que componen la retina en imágenes de OCT, diversos métodos han sido desarrollados. En el año 2011, Yazdan-panah et al. [33], propusieron un algoritmo semiautomático basado en los contornos Chan-Vese activo sin bordes con el propósito de segmentar 6 capas de la retina. Se trataba de un modelo Mumford-Shah multifásico y nivelado que incorporaba una forma previa de las capas de la retina basada en el conocimiento anatómico experto de dichas capas. Aplicaron su algoritmo sobre 80 imágenes OCT retinales de siete ratas y lograron un coeficiente de similitud de DICE promedio de 0.84 sobre todas las capas de la retina segmentadas.

En el año 2014, Srinivasan et al. presentaron un enfoque automático que utilizaba la eliminación de ruido basada en la teoría de *sparsity*, además utilizaba clasificadores de tipo *support vector machines*, teoría gráfica y programación dinámica para segmentar los límites de las capas retinianas en ratones de tipo salvaje (WT) y *rhodopsin knockout* ( $Rho(-/-)$ ) [34]. Para validar su método, se calculó el valor absoluto de la diferencia media de píxeles entre las segmentaciones automáticas y manuales para cada límite de capa. Sus resultados automáticos diferían de los del clasificador manual más experimentado en  $2,17 \mu m$  para el grupo WT y  $1,96 \mu m$  para el grupo  $Rho(-/-)$ .

Los trabajos mencionados anteriormente se basan en técnicas clásicas de procesamiento de imágenes; de hecho, los métodos de aprendizaje automático aún no se han utilizado ampliamente para segmentar las diferentes capas de la retina. Algunos métodos basados en *deep learning* se encuentran en la literatura para la segmentación de las capas retinianas pertenecientes a humanos. Un ejemplo de ello, es el trabajo realizado en el año 2017 por Leyuan Fang et al. en el que propusieron un nuevo método llamado (CNN-GS) para la segmentación automática de los 9 límites de capas de la retina humana en imágenes de OCT. Este método incluía un modelo de CNN para extraer características útiles de los límites específicos de la capa retiniana y, posteriormente, el entrenamiento de un clasificador para delinear las ocho capas. Por último, se llevaba a cabo una metodología de búsqueda gráfica sobre los mapas de probabilidad generados por la CNN para obtener los límites finales. Para evaluar la efectividad de su técnica, validaron su método en 60 volúmenes de OCT (2915 B-scan) de 20 ojos humanos [35].

Sin embargo, después de una exhaustiva búsqueda no ha sido posible encontrar ningún estudio previo que se haya centrado en la segmentación de las capas de la retina de roedores mediante la extracción automática de características utilizando técnicas de aprendizaje profundo.

Con respecto a la segmentación de las diferentes capas de la piel en imágenes de OCT, el trabajo realizado en la literatura es aún menos amplio que el llevado a cabo sobre las capas de la retina. Esto se debe a que el uso de la OCT en piel no está tan implantado como en retina.

En el año 2015, Annan Li et al. propusieron un método de segmentación de la epidermis que utilizaba tomografía de coherencia óptica de alta definición (HD-OCT). Los autores definían la segmentación de la epidermis en 3 pasos: preprocesamiento ponderado por mínimos cuadrados, detección de la superficie de la piel basada en gráficos y proyección integral local para la detección de la unión dermis-epidermis. El método propuesto se evaluó en un conjunto de datos que constaba de 5 volúmenes capturados por una máquina SkinTellR HD-OCT. Todas las muestras pertenecían a piel normal de los brazos con un tamaño de  $640 \times 200 \times 512$  [36].

El único trabajo previo de segmentación de imágenes OCT de piel en el que se implementan algoritmos basados en *deep learning*, fue realizado en 2019 por Calderon-Delgado et al. En este trabajo, se implementó una red neuronal Complementamente convolucional, en inglés conocida como Fully convolutional Network (FCN). El *ground truth* de las diferentes capas de la piel (dermis, epidermis, etc.) se obtuvo mediante métodos semiautomáticos y bajo la supervisión de investigadores experimentados. Como función de pérdidas se utilizó la validación cruzada categórica (del inglés - *categorical crossentropy* asignando diferentes pesos a cada capa para compensar las diferencias con respecto al área total. Sin embargo, estas imágenes presentaban estructuras complejas, como las células de Merkel o Langerhans, y folículos pilosos que no fueron segmentados [37]. Cabe destacar que algunos cánceres de piel y enfermedades como el acné se desarrollan en los folículos pilosos, siendo muy importante la segmentación de estas estructuras, tal y como afirman los especialistas que colaboran en el proyecto europeo GALAHAD.

Tras una extensa revisión en la literatura, se ha podido comprobar que no existe ningún trabajo previo en el que se hayan segmentado los folículos pilosos en imágenes OCT. Estas estructuras sí que han sido segmentadas en imágenes histopatológicas teñidas con hematoxilina y eosina, un pigmento que permite una mejora en la diferenciación de la epidermis y los folículos pilosos. Un ejemplo de ello es el trabajo realizado por Phillips et al. en el que se aplican técnicas de *machine learning* modernas para el diagnóstico del melanoma. En este proyecto se identifican tres tipos de tejidos diferentes (tumor, epidermis, dermis) en el que sí se incluye la segmentación de los folículos pilosos. Este trabajo está basado en el desarrollo de una *fully convolutional network* en el que se utilizaron parches de 512x512 de imágenes, pertenecientes a 50 imágenes de la biopsia de 49 pacientes [15].

Por tanto, se puede concluir que este proyecto es el primer estudio en el que se lleva a cabo una segmentación de la epidermis teniendo en cuenta los folículos pilosos en imágenes de OCT, simultáneamente a la detección de la dermis.

### 3.2.2 Principales capas que componen una red neuronal convolucional

Al igual que cualquier red neuronal tradicional, las redes convolucionales se componen de una capa de entrada, que en este caso es una imagen, de una de salida, que indica la probabilidad de pertenecer a una determinada categoría, y de una serie de capas ocultas, tal y como se puede observar en la Figura 3.5. Dentro de las capas ocultas se tienen capas convolucionales, capas de *batch normalization* o de “normalización”, capas de *pooling* o “diezmado”, capas de *flattening* o “aplanamiento” de los datos y capas *dense* o “densamente conectadas”.

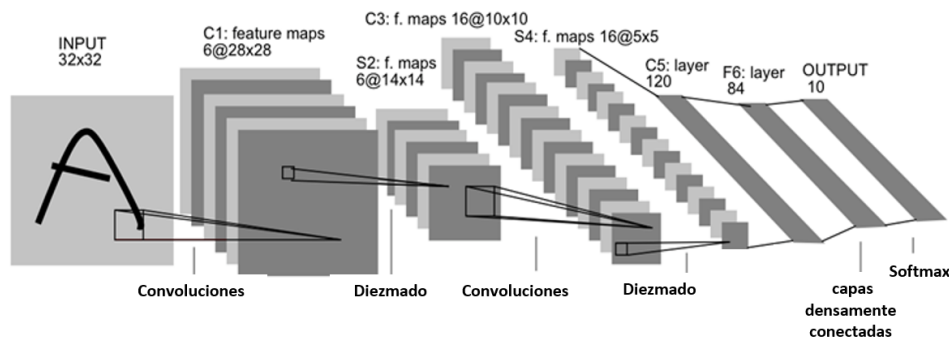


Figura 3.5: Ejemplo genérico de una red neuronal convolucional

A continuación, se explica la funcionalidad de cada una de estas capas siguiendo el orden con el que normalmente se disponen en las arquitecturas.

#### ■ Capas convolucionales

La capa de convolución es el bloque de construcción central de una red neuronal convolucional, ya que realiza la mayor parte del trabajo computacional más pesado y es la principal encargada de la extracción automática de características. Cada capa de convolución consiste en un conjunto de  $K$  filtros conocidos como kernels cuyos pesos son modificados durante el proceso de entrenamiento. Cada filtro es espacialmente pequeño, a lo largo de la anchura y la altura, ya que cuando se trata de entradas de alta dimensión, como imágenes, no es práctico conectar las neuronas de una capa a todas las neuronas del volumen anterior. Por lo tanto, en este caso, se conecta cada neurona únicamente a una región local del

volumen anterior. El alcance espacial de esta conectividad es un hiperparámetro<sup>2</sup> conocido como campo receptivo. Este equivale al tamaño del filtro y se describe como el conjunto de píxeles con los que el filtro realiza en cada paso la operación matemática conocida como correlación cruzada, cuya fórmula se expresa como:

$$y[m,n] = x[m,n] * h[m,n] = \sum_k \sum_l x[k,l]h[m+k,n+l] \quad (3.1)$$

Siendo  $x$  la imagen o mapa de activación de entrada,  $h$  el filtro aplicado e  $y$  la salida del filtro.

El *kernel* se va extendiendo por toda la dimensión del volumen de entrada en base a un hiperparámetro conocido como *stride*. Si el valor del *stride* es 1 como en la Figura 3.6, los filtros se moverán un píxel cada vez. Se debe tener en cuenta que en cada paso las conexiones son locales en el espacio, a lo largo de la anchura y la altura, pero siempre completas a lo largo de toda la profundidad del volumen de entrada.

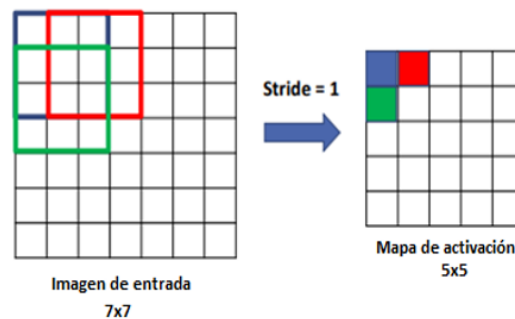


Figura 3.6: Filtro 3x3 recorriendo las dimensiones de la imagen con un *stride* = 1.

Como se puede observar en la Figura 3.6, tras la convolución, las dimensiones de la imagen de entrada se reducen, con la consecuente pérdida de información. Para preservar el volumen de entrada, es conveniente “rellenarlo” con ceros alrededor del borde, proceso conocido como *zero padding*. El tamaño de este “relleno” es un hiperparámetro que permite controlar el tamaño espacial de los volúmenes de salida. Si se pretende mantener las dimensiones del volumen de entrada ( $M \times N \times 3$ ), un *zero padding* de dimensiones  $N + (K_1 - 1) \times M + (K_2 - 1) \times 3$  debe ser computado antes de llevar a cabo la convolución, donde  $K_1 \times K_2$  es el tamaño del *kernel* y  $M, N$  el número de filas y columnas respectivamente de la imagen de entrada.

El resultado del proceso *zero padding* puede verse ejemplificado en la Figura 3.7.

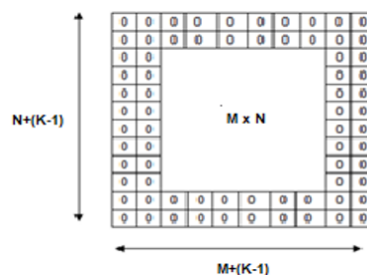


Figura 3.7: *Zero padding* utilizado para mantener las dimensiones de la imagen de entrada.

<sup>2</sup> Hiperparámetros: conjunto de parámetros modificables que permiten adaptar la arquitectura de una red neuronal.

A medida que se desliza el filtro sobre el ancho y alto del volumen de entrada, se produce un mapa de activación bidimensional que muestra las respuestas de ese filtro en cada posición espacial. Por tanto, cada filtro de cada capa convolucional producirá un mapa de activación bidimensional separado. Estos mapas se apilan a lo largo de la dimensión de profundidad  $Z$  para obtener el volumen de salida que pasará a capas posteriores, (véase Figura 3.8).

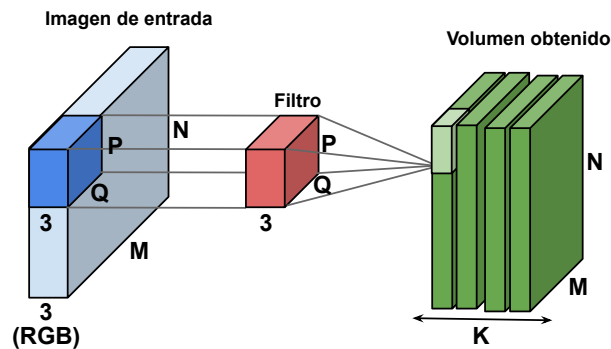


Figura 3.8: Volumen tras apilar los mapas de activación obtenidos en una capa convolucional

donde  $K$  hace referencia al número de filtros.

Los filtros pertenecientes a la primera capa de convolución tienen como entrada la imagen original y detectan características de bajo nivel como aristas y curvas. Cuando se pasa por otra capa de convolución, la salida de la primera capa convolucional se convierte en la entrada a la segunda. Ahora la entrada a esta capa no es la imagen original sino el mapa o mapas de activación que resultan de la primera capa. En este caso, la salida estará constituida por las activaciones que representan características de nivel superior, como semicírculos (combinación de una curva y un borde recto) o cuadrados (combinación de varios bordes rectos). A medida que se recorre la red y se atraviesan más capas convolucionales, se obtienen mapas de activación que representan características cada vez más complejas.

#### ■ *Capa de Batch normalization*

El entrenamiento de redes neuronales profundas puede llegar a ser un reto. Una posible razón de esta dificultad subyace en que la distribución de las entradas a las capas más profundas de la red puede cambiar después de cada *mini-batch*, cuando se actualizan los pesos de las capas iniciales [38]. Este cambio en la distribución de las entradas a las capas de la red recibe el nombre de “cambio de covariación interna”, en inglés *internal covariate shift*, y ralentiza el proceso de entrenamiento. Por otro lado, cuando la distribución de las entradas varía, también lo hace la salida de las neuronas. Esto produce que estas salidas fluctúen en las regiones saturables de algunas funciones de activación como la sigmoide. Una vez ocurre esto, las neuronas no pueden actualizar sus propios pesos ni pasar un gradiente de vuelta a las capas anteriores. Este fenómeno se conoce como *vanishing gradients* [39].

La aplicación de *batch normalization* mitiga el efecto de la variabilidad a las entradas de las capas. La idea es normalizar las entradas de manera que tengan media 0 y desviación estándar igual a 1. Además, con esto se consigue que la función de activación solo reciba entradas cercanas a cero asegurando el no *vanishing gradient* [39].

A continuación, se explican los pasos para conseguir que las entradas de cada función de activación tengan una distribución centrada en cero y con desviación la unidad.

En primer lugar, para obtener una distribución centrada en 0, se toma la entrada  $x$  y se le resta la media del lote o *minibatch*  $\mu_b$ , (ecuación 3.2).

$$z = x_i - \mu_B \quad (3.2)$$

Posteriormente, se divide  $z$  por la variación del *batch*  $\sigma_b$  y un número pequeño  $\varepsilon$  para evitar la división por cero  $\sigma_b + \varepsilon$ . Esto asegura que las distribuciones de entrada a las capas de activación tendrán como variación la unidad (ecuación 3.3).

$$\hat{x} = \frac{z}{\sigma_b + \varepsilon} \quad (3.3)$$

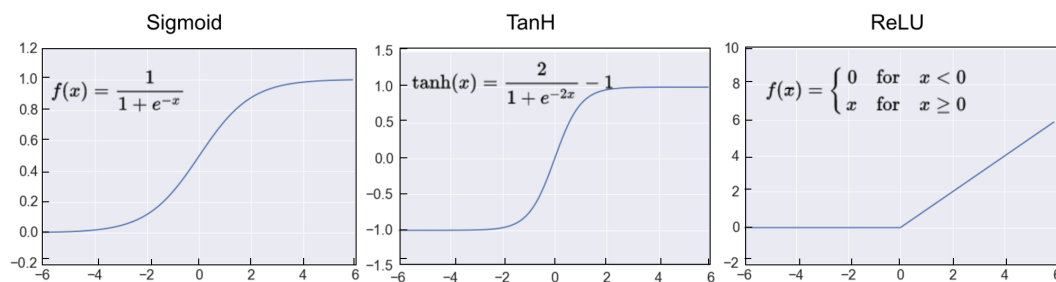
Por último, se pasa  $\hat{x}$  a través de una transformación lineal para escalar y cambiar la salida del *batch normalization*. Así se asegura que el efecto de normalización se mantiene a pesar de los cambios en la red durante la propagación hacia atrás.

$$y_i = \hat{x} * \lambda + \beta \quad (3.4)$$

donde  $\lambda$  y  $\beta$  son hiperparámetros a ajustar.

#### ■ *Capa de activación*

Los objetivos principales de estas capas son la introducción de no linealidades, ya que durante la convolución solo se han efectuado multiplicaciones y sumas, y la preservación de la dimensionalidad para mejorar la robustez. Existen diferentes funciones de activación como la sigmoide o la tangente hiperbólica que aparecen en la Figura 3.9. No obstante, la más usada en la actualidad es la RELU (*Rectified Lineal Unit*) que aplica la función  $f(x) = \max(0, x)$  cambiando las activaciones negativas a 0, proceso que acelera el entrenamiento sin comprometer la precisión [9].



**Figura 3.9:** Algunas de las funciones de activación existentes más importantes.

#### ■ *Capa de pooling*

Tras la aplicación de las funciones de activación, una red convolucional suele incluir capas de pooling. Estas capas reducen el tamaño espacial de representación con el objetivo de reducir el número de parámetros a ajustar y la complejidad en la computación de la red y, controlar así el *overfitting*. La más común es la función *maxpooling* que disminuye el tamaño del volumen de entrada tomando el máximo en un vecindario determinado. La forma más común es su utilización con filtros de tamaño  $F \times F$  con  $F = 2$



aplicados con *stride*  $S = 2$ , lo que descarta el 50% de las activaciones. La utilización de campos receptivos muy grandes ( $F$  muy grande) resulta demasiado destructivo. En la Figura 3.10 se puede observar un ejemplo de la actuación de esta capa para  $F = 2$  y  $S = 2$ .

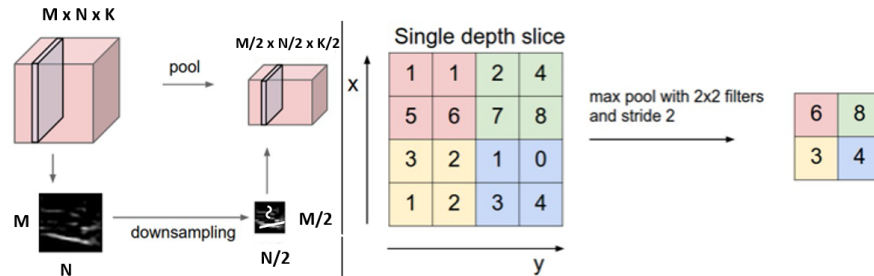


Figura 3.10: Aplicación de la capa de pooling con  $F = 2$  y  $S = 2$

■ **Capa completamente conectada**

Las últimas capas de la arquitectura de una CNN son las responsables de convertir el volumen de datos que se obtiene tras la salida de un bloque convolucional en un vector unidimensional que presenta el mismo número de elementos que el número de clases existentes.

Con el objetivo de conseguir lo expuesto en el párrafo anterior, después de aplicar una capa de *flatten*, que reorganiza los datos en una estructura unidimensional (vector), se aplica una capa conocida como *fully connected* que aplica una convolución 1D y reduce el número de activaciones al número de clases  $N_c$ . Posteriormente, tras utilizar una función de activación, comúnmente llamada *softmax*, se mapean las activaciones a las probabilidades finales. Esta función, ecuación 3.5, se encarga de calcular las probabilidades de cada clase dándoles un valor comprendido entre 0 y 1.

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{n=1}^{N_c} e^{z_n}} \quad (3.5)$$

Se cumple que la suma de las probabilidades de todas las clases es 1.

$$\sum \sigma(z)_j = 1 \quad (3.6)$$

A la salida de la capa *softmax* se obtiene un volumen con una profundidad igual al número de clases en el que se quiere clasificar la imagen,  $N$ . El valor localizado en el primer corte de este volumen, representa la probabilidad que tiene la imagen de pertenecer a la clase 1, el localizado en el corte 2, la probabilidad de pertenecer a la clase 2 y así hasta el número de clases. Por tanto, para obtener un solo valor que muestre la predicción dada por la red, se realiza la operación *max* a lo largo de la profundidad del volumen. Tras esta operación el valor de la predicción será igual a la clase que presenta una mayor probabilidad.

### 3.2.3 Entrenamiento de los algoritmos de segmentación de capas retinianas y epiteliales

Tras la explicación de la funcionalidad de las distintas capas por las que puede estar compuesta una CNN en la sección 3.2.2, se pueden plantear muchas preguntas tales como, ¿cómo saben los filtros de la primera capa convolucional buscar bordes y curvas?, ¿cómo se determinan los valores de los filtros de cada capa?, etc...

Antes de que se inicie la CNN, los pesos de los filtros de las capas convolucionales han sido inicializados aleatoriamente. Los filtros no saben buscar bordes y curvas, ni mucho menos características más complejas. Esto se asemeja al momento en el que una persona nace e inicialmente carece de conocimiento para reconocer cualquier objeto. Sin embargo, a medida que crece y aprende gracias a sus padres, por ejemplo, comienza a ser capaz de identificar determinados elementos. Con esta analogía, se pretende simular el comportamiento de una red neuronal, la cual es capaz de adquirir el conocimiento a partir de los datos etiquetados por un especialista. Esta idea de obtener una imagen y una etiqueta es el proceso de capacitación por el que pasan todas las CNNs, *backpropagation*, en el que se ajustan los valores de los filtros (o pesos) para poder obtener características relevantes de las imágenes. Este proceso se puede separar en 4 secciones distintas: paso hacia adelante o (*Forward pass*), cálculo de la función de pérdidas, paso hacia atrás o (*backward pass*) y actualización de los pesos.

#### *Paso hacia delante o Forward pass*

Durante el paso hacia adelante, el algoritmo toma una imagen de entrenamiento y la pasa a través de toda la red. En la primera iteración de entrenamiento, dado que todos los pesos o valores de los filtros se inicializan aleatoriamente, la salida no proporcionará ningún tipo de información. La red, con sus pesos iniciales, no puede buscar características de bajo nivel y, por lo tanto, no puede llegar a ninguna conclusión razonable con la que poder establecer la segmentación de la imagen.

#### *Función de pérdidas*

Tras el paso hacia delante se calcula el valor de la función de pérdidas, que controla el error que la red ha producido cuando se compara la segmentación de las imágenes de entrada con respecto a la imágenes etiquetadas.

Para el desarrollo de este proyecto, la función de pérdidas utilizada tanto para la segmentación de las imágenes de retina como de piel es una función logística ponderada de múltiples clases, *cross-entropy* ponderada. Esta proporciona una medida probabilística de similitud entre la predicción y el *ground truth*. Sea  $p_c(x)$  la probabilidad del píxel  $x$  de pertenecer a la clase  $c$  y  $g_c(x)$  el valor verdadero para el píxel dado por el *ground truth*, la función de pérdidas logística se puede definir como:

$$L(x) = - \sum_x \omega(x) g_c(x) \log(p_c(x)) \quad (3.7)$$

Los pesos  $\omega(x)$  se introducen en la función de pérdidas con dos objetivos: (i) compensar los efectos producidos por el carácter de desbalanceo de las clases, pues el diferente grosor de cada capa y la cantidad de píxeles que pertenecen al fondo propicia muestras de píxeles desequilibrados y (ii) aumentar la precisión de la segmentación en los límites de las capas [40]. Los píxeles cercanos a las regiones de transición del tejido son los casos más difíciles de segmentar correctamente debido a los límites difusos, el ruido *speckle* presente en las imágenes y la limitada resolución de la OCT.

Siendo  $f_c$  la frecuencia acumulada de la clase  $c$  en los datos de entrenamiento (es decir, la probabilidad de aparición),  $I$  el operador lógico (cuyo valor es 1 si se cumple la condición y 0 en el caso que no),  $M$  una imagen perteneciente al *ground truth* y  $\nabla$  el operador de gradiente 2D, el término  $\omega(x)$  se formula como:

$$\omega(x) = \sum_c I(M(x) = c) \cdot \frac{\mathbf{f}}{f_c} + \omega_0 \cdot I(|\nabla M(x)| > 0) \quad (3.8)$$

donde  $\mathbf{f} = [f_1, f_2, \dots, f_c]$  es un vector que contiene todas las frecuencias.

El primer término modela el equilibrio de frecuencias medias y compensa el problema de desequilibrio de clases al darle más peso a las clases con baja probabilidad [41]. El segundo término establece un mayor peso en las regiones de los límites anatómicos para enfatizar la segmentación correcta de los contornos [42].  $\omega_0$  equilibra ambos términos.

Como se puede imaginar, las pérdidas en las primeras fases de entrenamiento son extremadamente altas. El objetivo es llegar a un punto donde la etiqueta predicha sea la misma que la etiqueta de entrenamiento, lo que significa que la red ha cumplido con su predicción correcta. Para conseguir este objetivo, es necesario minimizar las pérdidas obtenidas averiguando qué entradas (pesos de los filtros) contribuyeron en mayor proporción a la pérdida (o error) de la red. Por tanto, el problema de aprendizaje en las redes neuronales se formula en términos de la minimización de la función de pérdidas asociada [43].

Como se ha podido observar en la ecuación 3.7, la función de pérdidas o error es una función no lineal, por lo que no se disponen de algoritmos sencillos y exactos para encontrar sus mínimos. En consecuencia, se tendrá que realizar una búsqueda a través del espacio de parámetros para encontrar aquellos que proporcionen, de forma aproximada, el mínimo error en la función de pérdidas [43].

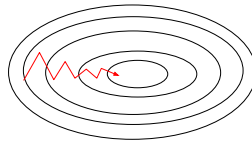
#### *Paso hacia atrás o backward pass y actualización de pesos*

Tras obtener el error producido, el próximo paso es realizar un recorrido hacia atrás a través de la red para ajustar los pesos a fin de que el valor de la función de pérdidas disminuya. Una vez minimizada esta función, se procede con la actualización de los pesos.

El algoritmo utilizado en este proyecto para encontrar los mínimos de la función de pérdidas se denomina descenso por gradiente estocástico, del inglés - Stochastic Gradient Descent (SGD). Este método es un caso especial del método de descenso por gradiente. En este último, se debe calcular el coste y el gradiente para todo el conjunto de entrenamiento, tarea que puede ser muy lenta y a veces difícil de realizar en una sola máquina si el conjunto de datos es demasiado grande [44].

El algoritmo SGD soluciona el problema anteriormente citado, ya que en cada época el entrenamiento se compone de un conjunto de  $N$  iteraciones en las que se calcula el gradiente solo para unos pocos datos de entrenamiento o *minibatch* [44] y se computa la actualización de pesos correspondiente. El número de iteraciones que compone una época se define como  $N = \frac{\text{Cantidad de datos de entrenamiento}}{\text{Longitud del minibatch}}$ .

En este proyecto, se ha utilizado un *minibatch* de 8 muestras. Como en cada iteración no se está usando todo el conjunto de entrenamiento, cada paso dado no tiene porqué ser en la dirección correcta. Esto significa, que es posible que se necesiten más pasos para converger (ver Figura 3.11), pero estos computacionalmente son menos costosos [45].



**Figura 3.11:** Método de descenso por gradiente estocástico

En el método SDG, para realizar la actualización de un parámetro o peso  $w_{i+1}$  a partir de  $w_i$ , que son los pesos calculados en la iteración  $i$  utilizando los gradientes de la iteración  $i - 1$ , se traslada ese punto en la dirección de entrenamiento  $d_i = g_i$  siendo  $g_i = \nabla L(w_i)$ , por tanto:

$$W_{i+1} = W_i - \alpha * \nabla L(W_i; x^{(k)}, y^{(k)}) \quad (3.9)$$

con el par  $(x^{(k)}, y^{(k)})$  pertenecientes al conjunto de entrenamiento.

El parámetro  $\alpha$  se denomina tasa de entrenamiento. En este caso, tanto para piel como para retina, se establece inicialmente a 0.001 y se reduce en un orden de magnitud cada 20 épocas de entrenamiento. Este parámetro determina el tamaño del paso en la dirección del gradiente negativo. Una alta tasa de aprendizaje significa que se toman pasos más grandes en las actualizaciones de peso y, por lo tanto, el tiempo para que el modelo converja es menor. Sin embargo, una velocidad de aprendizaje demasiado elevada podría dar como resultado saltos demasiado grandes y no lo suficientemente precisos para alcanzar el punto óptimo, quedándose el algoritmo estancado en un mínimo local o diverger.

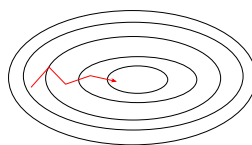
El SGD tiene problemas para navegar por áreas en las que la curva se inclina más hacia una dimensión que otra. Estas zonas son comunes cerca de los óptimos locales, y son abundantes en redes neuronales profundas. En estas áreas, el SGD oscila a través de sus pendientes, lo que hace que el progreso sea muy lento [45]. El *momentum* es un método que ayuda a acelerar el SGD en la dirección correcta. Este efecto se consigue añadiendo una fracción  $\gamma$  del vector de pesos de la iteración anterior al vector de pesos en la iteración actual. Este término se define como:

$$M_i = \gamma * W_{i-1} + \alpha \nabla L(W_i; x^{(k)}, y^{(k)}) \quad (3.10)$$

Después, usando el término  $M_i$ , se modifican los pesos de la siguiente forma:

$$W_{i+1} = W_i + M_i \quad (3.11)$$

Como se puede deducir de su expresión matemática, el parámetro *momentum* se incrementará para aquellos gradientes que apunten hacia una misma dirección de forma consistente. En caso contrario, se reducirá. Esto disminuye las oscilaciones y acelera la convergencia, (véase Figura 3.12).



**Figura 3.12:** Método de descenso por gradiente estocástico corregido por el parámetro *momentum*.

En este caso, se ha elegido un  $\gamma$  de 0.9 para los dos tipos de entrenamientos llevados a cabo en este proyecto.

Una vez se completan todos los pasos que componen el *backpropagation* sobre los datos del conjunto total de entrenamiento se da por finalizada una época. En este proyecto, tanto para el entrenamiento con imágenes retinianas como de piel, se han entrenado los algoritmos durante 60 épocas. Para cada época de entrenamiento se obtiene un modelo diferente que sustituirá al anteriormente guardado si sus pérdidas en validación son menores que en la anterior época.

### 3.2.4 Segmentación de las capas de la retina y piel mediante encoder-decoder

La tarea de segmentación de las capas retinianas y epiteliales propuestas en este proyecto puede interpretarse como un problema de clasificación. Dado un píxel de una imagen de OCT  $p = f(x, y)$ , el objetivo es asignarlo a su etiqueta correspondiente en el espacio  $L = 1, \dots, N$  siendo  $N$  el número de clases posibles [46]. Particularmente, en los casos propuestos en este proyecto,  $N = 4$  para la segmentación de las capas retinianas y  $N = 3$  para el caso de la piel.

Para conseguir el objetivo de segmentación propuesto, se diseñan dos tipos de redes neuronales convolucionales con una estructura *encoder-decoder* construidas sobre ReLaynet [46]. En este tipo de redes, la arquitectura se divide en dos partes, una parte de *encoder* y otra de *decoder*.

El *decoder* está basado en una arquitectura generalmente de la misma estructura que la que tiene la parte de *encoder* pero en orientación opuesta. Esta parte se encarga de recoger las características extraídas por el *encoder* para construir la imagen de salida con la mayor precisión posible y con la misma dimensión espacial que la imagen de entrada. Normalmente se establecen conexiones *skip connections* ente la parte del *encoder* y el *decoder* para transmitir las representaciones de las características de los bloques codificadores a los bloques decodificadores emparejados a través de capas de concatenación. Esto ayuda a recuperar los detalles del objeto de estudio [47], [48].

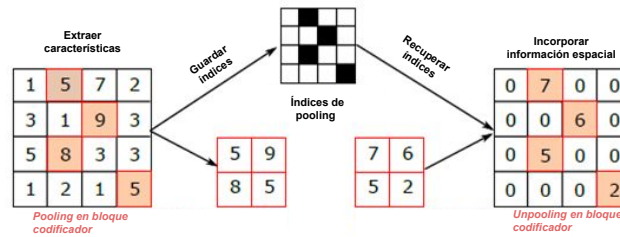
A continuación, se expone la composición de cada bloque *encoder* y *decoder* propuesto, tanto para la segmentación de las capas retinianas como para la segmentación de la capa epitelial.

#### ■ *Encoder*

Cada bloque de *encoder* está compuesto por cuatro capas, en secuencia: capa de convolución, capa de *batch normalization*, capa de activación ReLU y capa de *max pooling*. Para capturar las transiciones entre las diferentes capas, cada capa de convolución consta de 64 filtros. En este caso, *kernels* rectangulares de tamaño  $7 \times 3$ . Además, se utiliza *zero padding* para preservar las dimensiones espaciales de los mapas de activación después de cada capa convolucional, tal y como se ha explicado en secciones anteriores. Posteriormente, se aplica una técnica de *batch normalization* para evitar el *overfitting* durante el procedimiento de entrenamiento [49]. En cuanto a la función de activación ReLU, esta permite introducir no linealidades, lo que conlleva un entrenamiento mucho más rápido. La operación de *max-pooling* se introduce en la última capa de cada bloque *encoder* para condensar la información reduciendo las dimensiones espaciales. Esta reducción depende del tamaño del kernel y del paso (*stride*) utilizado. En este caso, las dimensiones se reducen a la mitad, ya que se utiliza un *kernel* con *stride*  $S = 1$ . Además, para mantener la consistencia espacial, lo que es de gran importancia en lo que segmentación semántica se refiere, los índices de *pooling* derivados de esta operación se transfieren a la capa de *unpooling* correspondiente en el bloque del *decoder*.

■ **Decoder**

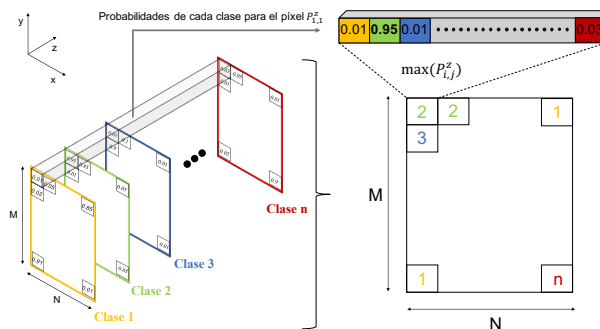
Cada bloque de *decoder* consta de cinco capas, en secuencia: capa de *unpooling*, capa de concatenación, capa de convolución, capa de *batch normalization* y función de activación ReLU. La capa *unpooling* interpola los mapas de características del bloque decodificador anterior a una resolución más fina mediante el uso de los índices de *pooling* que han sido guardados y pertenecen al bloque codificador emparejado. El mapa de características obtenido tras esta capa es ejemplificado en la Figura 3.13.



**Figura 3.13:** Proceso para incorporar la información espacial recuperando los índices que se han mantenido tras aplicar la capa de *pooling*.

En la Figura 3.13 se puede observar que únicamente tienen un valor distinto de 0 aquellos índices que coinciden con las posiciones de los píxeles mantenidos tras la capa de *max-pooling* del proceso de *encoder*. Este proceso mejora enormemente la preservación de la información espacial. Después de este paso se realiza una concatenación de los mapas de características interpolados con los mapas de características de salida correspondientes al bloque codificador, *encoder*, emparejado para enriquecer la información y evitar problemas de degradación. Finalmente, las capas convolucionales, las de *batch normalization* y la función de activación ReLU se aplican al mapa de características concatenado.

El bloque *decoder* final, bloque de clasificación, está compuesto por una capa convolucional con un *kernel* 1x1 y una función de activación *softmax*. A la salida de esta capa se obtiene un volumen con una profundidad igual al número de clases en el que se quiere segmentar la imagen  $N$ . El valor de cada píxel localizado en el primer corte de este volumen representa la probabilidad que tiene de pertenecer a la clase 1, el localizado en el corte 2, la probabilidad de pertenecer a la clase 2 y así hasta el número de total clases. Por tanto, para obtener una sola imagen que muestre la predicción dada por la red, se realiza la operación *max* a lo largo de la profundidad del volumen. Tras esta operación el valor de cada píxel será igual a la clase que presenta una mayor probabilidad. Este proceso puede verse ejemplificado en la Figura 3.14.



**Figura 3.14:** Operación realizada tras la función de activación *softmax*

En la Figura 3.15 se muestra la arquitectura utilizada para llevar a cabo el objetivo de segmentación propuesto en imágenes de retina, es decir, la obtención de los límites de las capas ILM, IPL-INL e IS-OS descritas en la sección 2.1.1 y visualizadas en imágenes de OCT en la sección 2.3.3.

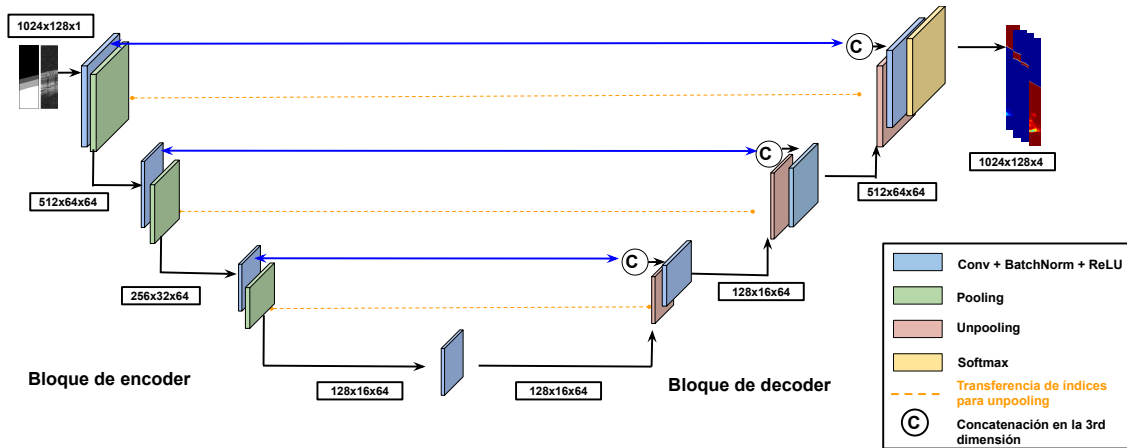


Figura 3.15: Arquitectura utilizada para segmentar las diferentes capas de la retina.

Como se puede observar, se utilizan tres bloques de *encoder* (o extracción de características) y tres de *decoder*, con la composición citada en párrafos anteriores.

Con respecto a la arquitectura utilizada para segmentar las fronteras de la capa epitelial de la piel representada en la sección 2.3.3, se expone que esta es idéntica a la que se utiliza para extraer las capas retinianas con la salvedad de que en la parte de *encoder* y *decoder* se le añade un bloque. Este efecto se puede observar en la Figura 3.16. Esto se debe a que para poder optimizar la segmentación en las imágenes de piel es necesario crear una red más profunda, puesto que los límites en estas imágenes, sobre todo en la unión dermo-epidérmica, son más difusos que en las imágenes de retina, por lo que la extracción de características es más compleja.

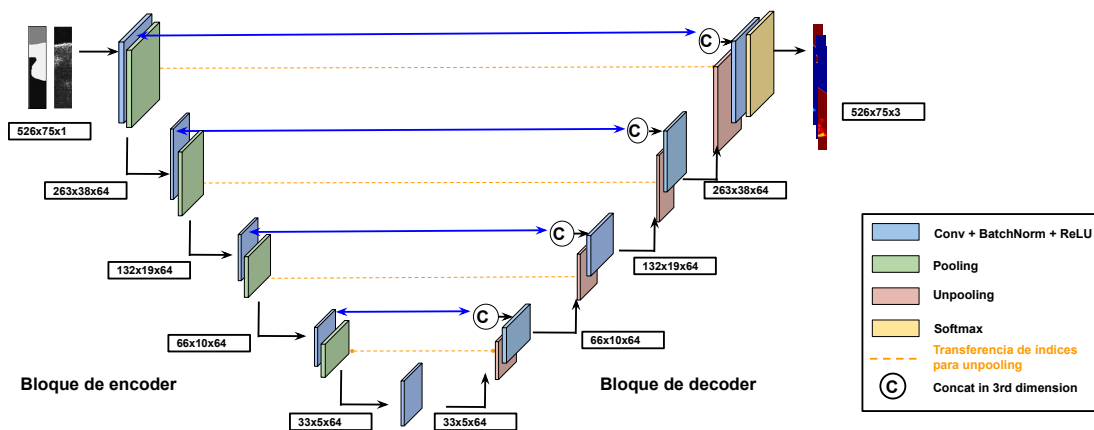


Figura 3.16: Arquitectura utilizada para segmentar las diferentes capas de la piel.

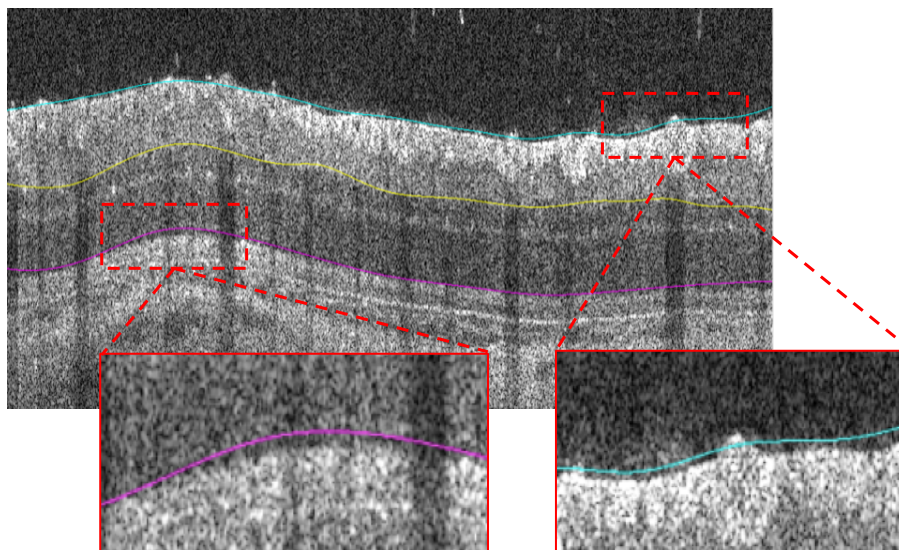
### 3.2.5 Generación del ground truth

Como se ha explicado a lo largo de esta memoria, los objetivos del presente TFG residen en obtener, por un lado, las capas más importantes que componen la retina y, por otro lado, los límites de la capa epitelial.

En la sección 3.2.4 se ha detallado que, para lograr estos objetivos, se utilizan dos arquitecturas basadas en redes neuronales convolucionales. Este tipo de algoritmos están basados en técnicas de aprendizaje supervisado. En ellos se utiliza un conjunto de datos de entrenamiento (datos de entrada y sus etiquetas correspondientes) a fin de generar un modelo que realice predicciones acerca de las etiquetas para un nuevo conjunto de datos [50]. Es por ello que, para entrenar estos algoritmos es necesario generar las etiquetas adecuadas. Para ello, se deben obtener las fronteras de las regiones de interés, es decir, se debe realizar una segmentación manual para, posteriormente, poder etiquetar cada región. A continuación, se explican con detalle los pasos a seguir para obtener el *ground truth*, tanto en las imágenes de la retina como en las imágenes de piel.

#### *Ground truth en imágenes de retina*

Como se ha descrito en la sección 1.2, este trabajo se enmarca dentro del proyecto GALAHAD en el que participan diversas entidades europeas. Uno de ellos, RegH<sup>3</sup>, que aportó las dos entregas de la base de datos de roedores, también proporcionó una primera aproximación del ground truth. Este puede ser visualizado de forma cualitativa en la Figura 3.17.



**Figura 3.17:** *Ground truth* proporcionado por el miembro RegH perteneciente al proyecto GALAHAD.

Como se puede observar en esta figura, la segmentación inicial presenta ciertas imperfecciones, por lo que es necesario modificarla. Para llevar a cabo la corrección de la segmentación aportada y obtener el *ground truth* definitivo, se ha desarrollado para este proyecto una interfaz gráfica de usuario (GUI, del inglés – Graphic User Interface) en el entorno de Matlab. Como se puede observar en la Figura 3.18,

---

<sup>3</sup>REGH: acrónimo de Region Hovedstaden (Rigshospitalet-Glostrup), hospital procedente de Dinamarca que aportó las imágenes de la base de datos de roedores



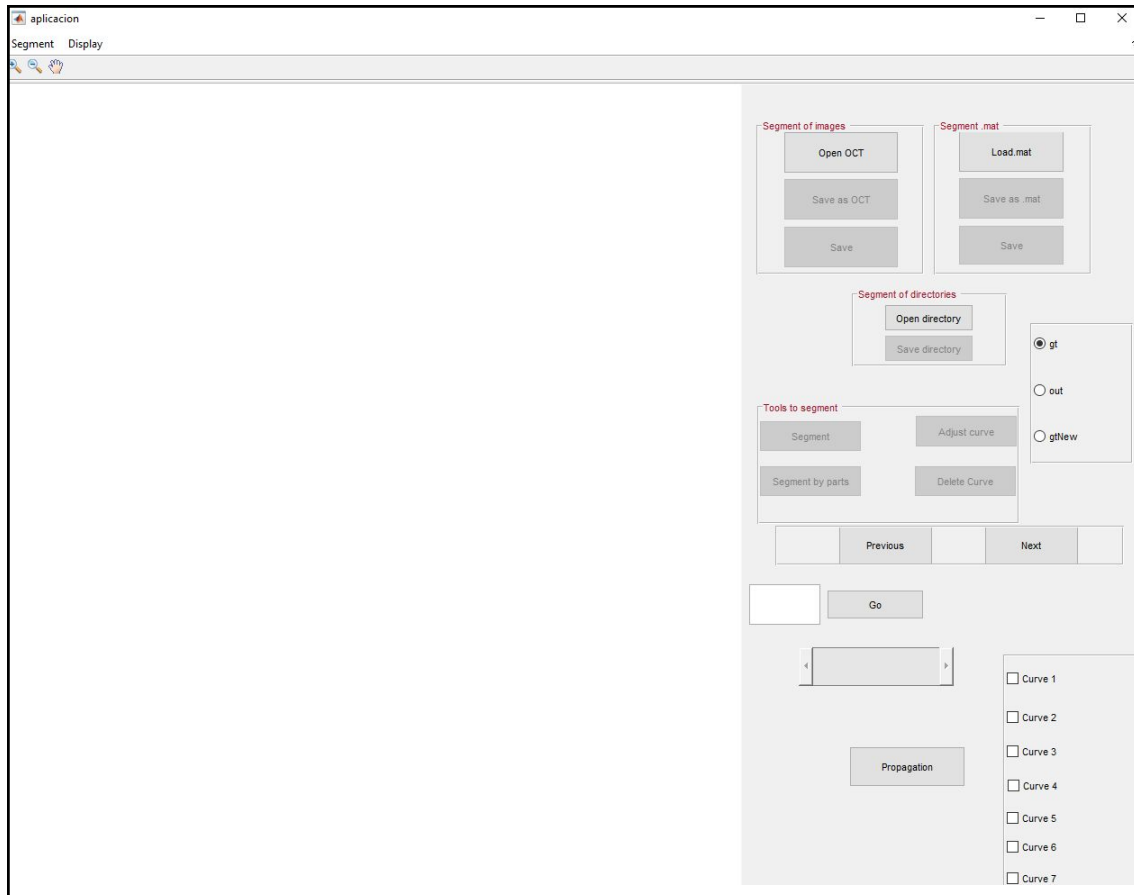


Figura 3.18: Aplicación desarrollada para generar el *ground truth* definitivo

la ventana principal de la aplicación es intuitiva para que cualquier persona sin conocimiento de programación, por ejemplo un especialista sanitario, pueda llevar a cabo la modificación o la creación de una nueva segmentación.

Esta parte principal de la aplicación permite cargar directorios de imágenes, archivos con extensión .OCT y archivos .mat. Tras cargar el archivo correspondiente, la aplicación permite realizar una segmentación desde cero, borrar o modificar la segmentación existente. En este caso, para obtener el **ground truth** definitivo se tuvo que usar la funcionalidad que aporta el botón *adjust curve* y que permite modificar las curvas existentes.

Una vez se carga el archivo .mat donde se tienen guardadas las imágenes, se pulsa sobre el botón *adjust curve*. Tras pulsar este botón aparecen 32 puntos de cada curva. Con el botón izquierdo del ratón, se pulsa sobre la curva que se pretende modificar, apareciendo en rojo el punto que se está modificando en cada momento. Con las flechas del teclado (izquierda y derecha) se va cambiando el punto que se modifica de píxel en píxel, mientras que si se tiene pulsada la tecla *shift* y las mismas flechas, los puntos se desplazan de tres en tres píxeles. Con las flechas (arriba y abajo) se aumenta o disminuye respectivamente la altura de los puntos en un píxel y, si se mantiene pulsada la tecla *shift* y las mismas flechas, se modifica en saltos de 5 píxeles. Cuando se acaba de modificar la curva, con la tecla *intro* se interpola la misma, de forma que el número de puntos de cada curva es igual al número de columnas presentes en la imagen. Para ello, se utiliza la función *spline* de Matlab. Esta función realiza una interpolación utilizando un spline cúbico que, a grandes rasgos, es una curva diferenciable definida en porciones mediante polinomios de

grado 3. Posteriormente, se sobrescribe la modificación sobre la anterior segmentación y se obtiene el *ground truth* definitivo. Este resultado se puede visualizar pulsando sobre el botón *gtnew* presente en el conjunto de *radio button*.

En la Figura 3.19 se ejemplifica el modo de actuar para obtener el *ground truth* definitivo.

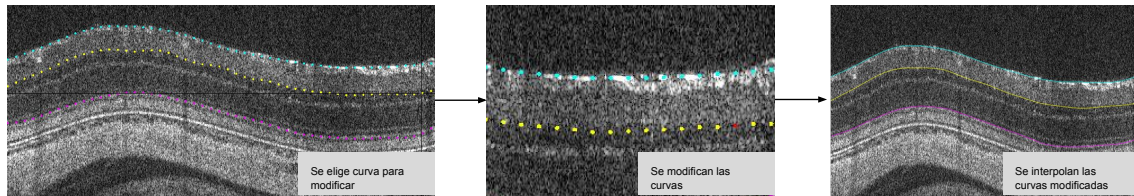


Figura 3.19: Modo de actuación para obtener el *ground truth* definitivo.

Tras hacer las modificaciones pertinentes, en el submenú *display* de la aplicación, se pueden observar las modificaciones realizadas en cada una de las curvas. Por ejemplo, en la Figura 3.20 se puede visualizar la segmentación sin modificaciones en rojo, mientras que las modificaciones que se han realizado para obtener el *ground truth* definitivo aparecen en amarillo.

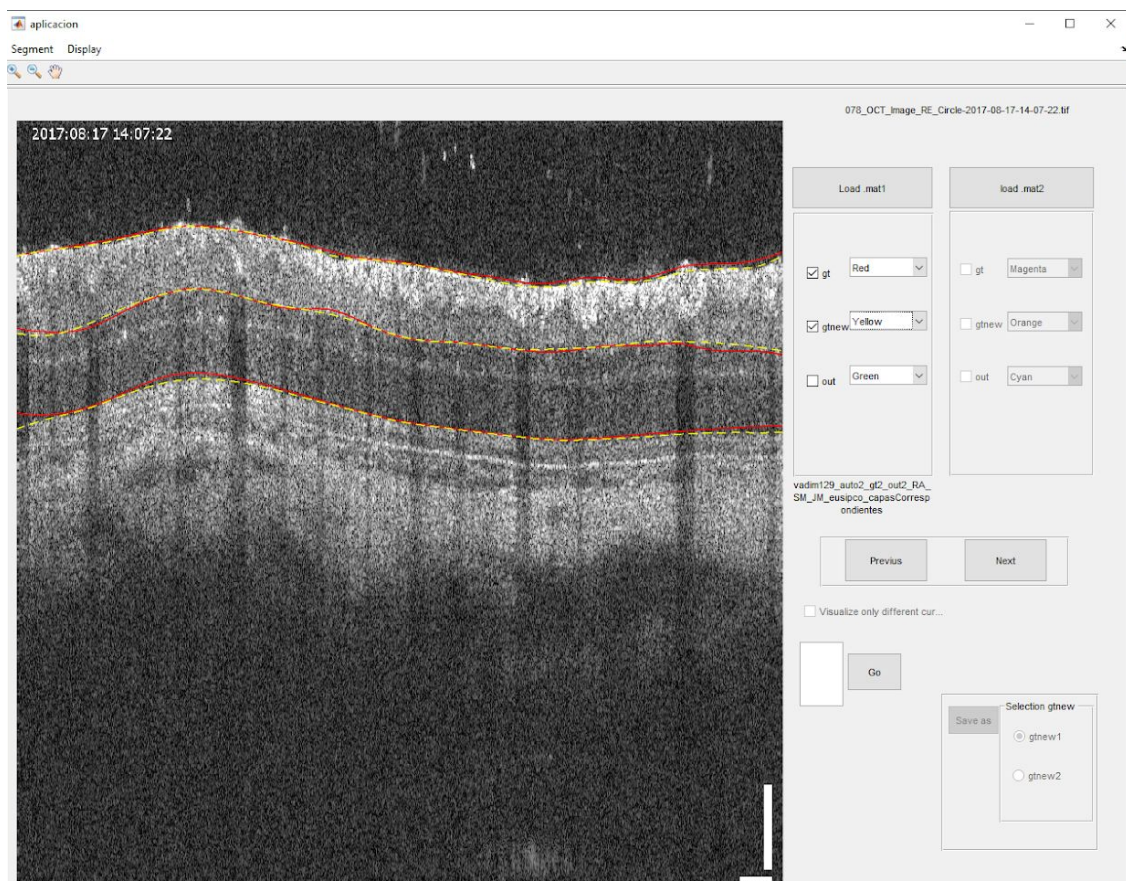
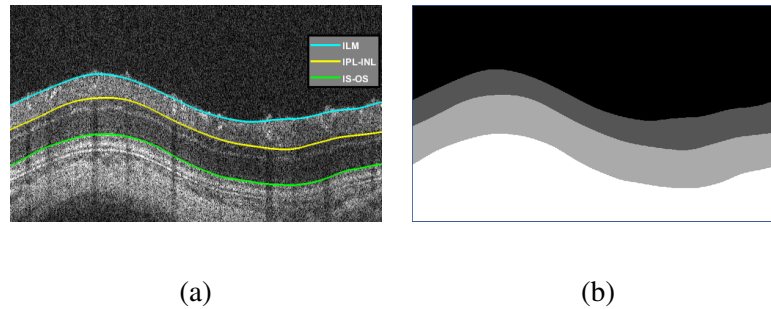


Figura 3.20: Pasos efectuados para obtener el *ground truth* definitivo.

Tras segmentar las fronteras de aquellas regiones de interés, es necesario asignar a los píxeles que se encuentran dentro de ellas una clase diferente y crear una imagen de *layers*. En la Figura 3.21 se puede apreciar un ejemplo de ello.



**Figura 3.21:** Obtención de la imagen de *layers* a partir de las curvas segmentadas: a) Imagen de *boundaries* b) Imagen de *layers*. (ILM: Membrana limitante externa, IPL: Capa plexiforme externa, INL: Capa nuclear interna, IS: Segmento interno, OS: Segmento externo).

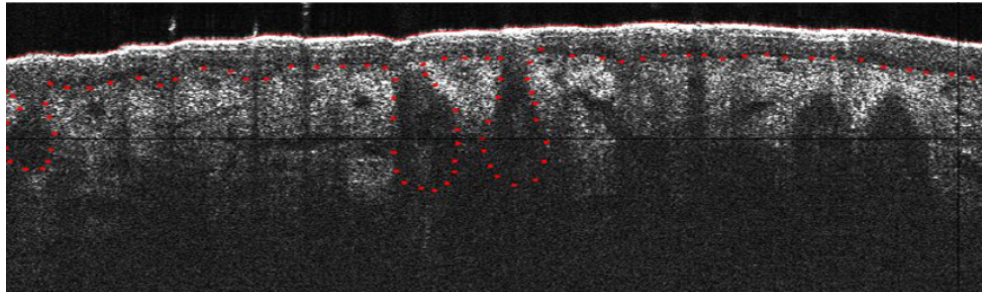
Como se puede observar en la Figura 3.21 se dispone de cuatro clases diferentes: límite superior, capa RNFL+GCL+IPL, capa INL+OPL+ONL+IS-OS y límite inferior con el valor de 0, 1, 2, 3 y 4, respectivamente. Esta es la imagen etiquetada que el algoritmo utilizará como etiqueta para entrenar el modelo.

### ***Ground truth en imágenes de piel***

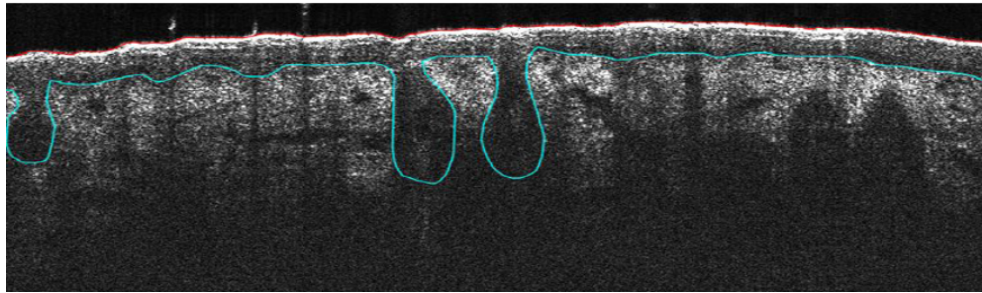
Para la realización del *ground truth* en las imágenes de piel, no se contaba con ninguna segmentación previa. Por ello, fue necesario realizar una segmentación desde cero, haciendo uso del botón *segment* de la aplicación mostrada en la Figura 3.18. Tras pulsar sobre este botón, con el cursor izquierdo del ratón se pintan los puntos que la curva debería seguir. A continuación, se pulsa sobre el cursor derecho del ratón y se interpolan los puntos para generar la curva correspondiente. Sin embargo, en este caso no fue posible realizar directamente la interpolación con la función *spline* de Matlab. Esta función impone que las abscisas de los puntos en los que se quiere realizar la interpolación  $P_k(x_k, y_k)$ , siendo  $y_k = f(x_k)$ , deben estar ordenados, es decir,  $x_0 < x_1 < x_2 < \dots < x_n$  [51].

Como se puede observar en la Figura 3.22, las abscisas correspondientes a la segunda curva (azul) presente en las imágenes de piel, unión dermo-epidermis, están desordenadas<sup>4</sup> debido a la presencia de los folículos pilosos. Para hacer frente a este problema, se llevó a cabo una función que realiza una interpolación por partes. Para ello, en primer lugar, se calcula la derivada de los puntos que componen la curva. Posteriormente, se detectan las posiciones de los puntos que presentan una derivada negativa y se interpola por partes entre estas posiciones. Por último, se unen todas las interpolaciones realizadas en una sola curva.

<sup>4</sup>Los puntos se encuentran desordenados debido a que por la forma elipsoidal de los folículos algunos valores de  $x$  tienen un menor valor que los anteriores



(a)



(b)

**Figura 3.22:** Segmentación de las imágenes de piel. En la Figura a) aparecen los puntos marcados por el usuario para la generación de la curva. En la figura b) aparece el resultado obtenido tras la interpolación de los puntos, región superior de la epidermis (curva roja), unión dermo-epidérmica (curva azul).

Cabe destacar que la segmentación manual fue realizada por la autora de esta memoria, en base al conocimiento adquirido en asignaturas de histología en la carrera de Ingeniería Biomédica. No obstante, es importante remarcar que dicha segmentación manual fue supervisada por expertos en la materia, que ofrecían un *feed-back* de los cambios que debían realizarse en cada momento.

Tras realizar la segmentación de las fronteras de la epidermis, se procedió de la misma forma que con las imágenes de retina para obtener las imágenes de *layers* como la que se puede observar en la Figura 3.23.



**Figura 3.23:** Imagen de *layers* obtenida a partir de las coordenadas de las diferentes curvas segmentadas.

### 3.2.6 Partición de los datos

Tras la generación del *ground truth*, para que pueda llevarse a cabo el proceso de entrenamiento de una red neuronal, es fundamental determinar qué muestras del conjunto de datos se utilizarán para el entrenamiento. Por tanto, para garantizar la robustez de los modelos implementados, es necesario realizar una partición de los datos para entrenar y testear los modelos de segmentación propuestos. Si se entrenan los algoritmos con los mismos datos que posteriormente se utilizan para testear, los resultados obtenidos serían inmejorables, pero engañosos, ya que los modelos obtendrían peores resultados si predijesen datos que nunca han sido vistos por la red.

Para evitar resultados sesgados debido a una partición específica de la base de datos en subconjuntos de entrenamiento y test y, poder obtener la segmentación de todas las imágenes de la base de datos, se lleva a cabo una técnica conocida como *k-fold cross validation*. Esta metodología queda ejemplificada en la Figura 3.24.

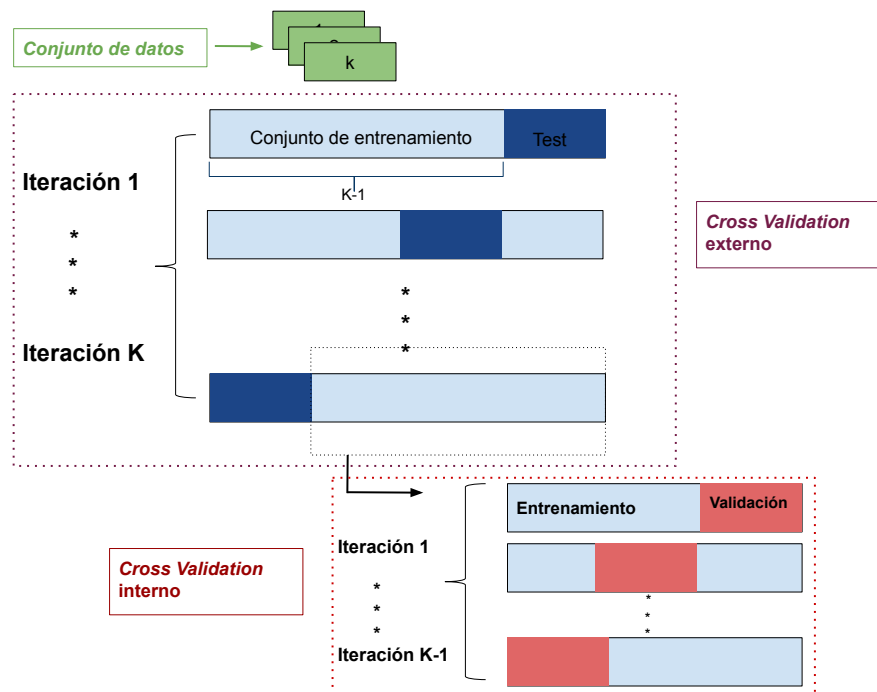


Figura 3.24: Metodología *K-fold cross validation*.

En dicha figura, el conjunto de datos se divide en  $k$  subconjuntos (de la misma cantidad de muestras aproximadamente) para evitar el problema de desbalanceo entre clases. De esta forma,  $(k - 1)$  subconjuntos de datos se utilizan para entrenar y validar los modelos, mientras que el último conjunto se usa para testearlos. El proceso se repite  $k$  veces hasta asegurar que todos los subconjuntos han sido utilizados para testear y para entrenar, pero en iteraciones diferentes. Además se implementa una validación cruzada interna, utilizando las imágenes de un subconjunto de entrenamiento diferente en cada iteración interna como conjunto de validación. Como resultado de este proceso, se obtienen  $(k - 1) k$  combinaciones diferentes.

Específicamente, para las imágenes de retina se crean  $k=8$  particiones. Las imágenes pertenecientes a las ratas 1-7 componen siete particiones individuales (aproximadamente 16 imágenes en cada partición), mientras que las imágenes de las últimas tres ratas (4, 8 y 4 imágenes, respectivamente) se agrupan

formando el octavo conjunto para evitar problemas de desequilibrio de las clases. En consecuencia, en el proceso de entrenamiento, se utilizan 7 conjuntos diferentes en cada iteración externa (88% de los datos), mientras que la partición restante se emplea para testear el rendimiento del modelo. Con respecto a la validación interna, conviene señalar que se valida cada vez con uno de los 7 conjuntos de los datos restantes.

En el caso de las imágenes de piel, cada paciente compuesto por 30 imágenes se incluye dentro de un subconjunto distinto, generando un total de  $k=9$  particiones o subconjuntos. Por tanto, en cada iteración externa se utilizan 8 subconjuntos para la fase de entrenamiento (90% de los datos) y uno para testeo (10% de los datos). Con respecto a la validación interna se valida cada vez con uno de los 8 subconjuntos de los datos restantes.

Por otra parte, debido a que las imágenes utilizadas para el entrenamiento tienen grandes dimensiones, en retina  $1024 \times 1024$  y en piel  $526 \times 975$ , se aplica una metodología de aprendizaje por parches para evitar problemas de memoria. En este caso, se emplean parches de las imágenes de  $1024 \times 128$  en retina y  $526 \times 75$  en piel. Además, en ambos casos se aumentan los datos mediante la introducción de transformaciones geométricas aleatorias como recortes, giros horizontales, rotaciones y traslaciones. Esta técnica es conocida como *data augmentation* y permite aumentar el conjunto de datos creando imágenes sintéticas similares a las originales (con la etiqueta correspondiente), sin necesidad de aumentar el número de imágenes anotadas.

### 3.2.7 Testeo de los algoritmos de segmentación de capas retinianas y epiteliales

Tras entrenar los diferentes modelos, es necesario realizar una predicción para cada una de las imágenes pertenecientes a la base de datos. Como se ha comentado en la sección 3.2.6, se ha hecho uso de una técnica de *cross validation* tanto externa como interna obteniendo un conjunto de  $k \times (k - 1)$  combinaciones de test y validación, siendo  $k$  el número de subconjuntos en los que se han dividido los datos. Es decir, para cada subconjunto de test, se emplean  $(k - 1)$  modelos para predecir las imágenes de ese subconjunto, uno por cada subconjunto de validación diferente utilizado en el entrenamiento.

Para obtener la predicción de los diferentes subconjuntos de test, se calcula la predicción de las imágenes de test para cada uno de los  $k - 1$  modelos  $P_j$  (referente a los subconjuntos con los que se ha validado), computándose la predicción final para cada subconjunto de test  $P_i$  de la siguiente forma:

$$P_i = \frac{1}{K-1} \sum_{j=1}^K P_j \quad \forall j \neq i \quad (3.12)$$

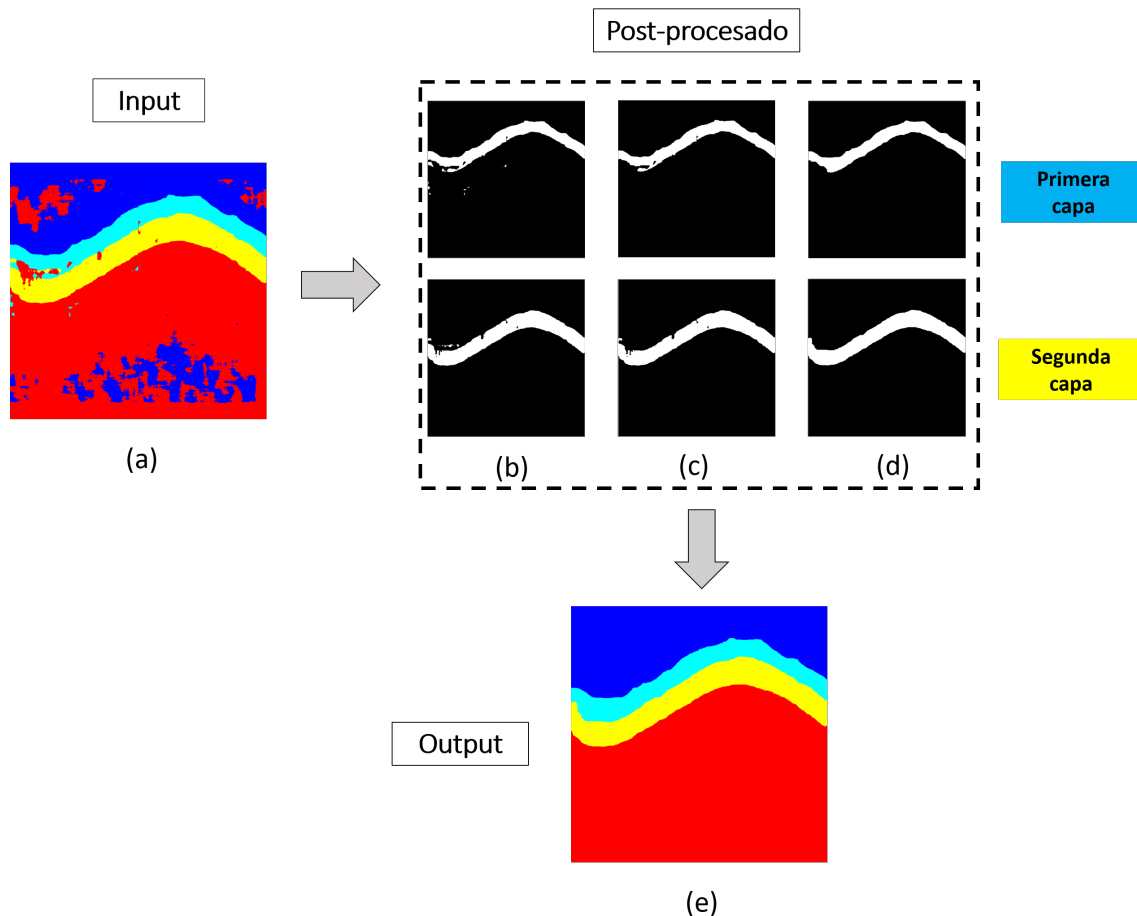
La predicción tendrá una profundidad igual al número de clases, por lo que se calcula el máximo en la dirección de la profundidad, obteniendo para cada píxel la clase cuya probabilidad de pertenencia es mayor.

Además, cabe destacar que para las imágenes de retina se ha contado con una segunda entrega de la base de datos diferente a aquella con la que se ha realizado el entrenamiento. Haciendo uso de esta, se ha realizado un segundo testeo con los modelos entrenados a partir de la primera entrega de la base de datos. Con esto se pretende comprobar la capacidad del algoritmo desarrollado a la hora de predecir nuevas imágenes pertenecientes al mismo tipo de sujeto y al mismo órgano de estudio.

### 3.2.8 Post-procesado

#### Post-procesado en imágenes de retina

Para la completa definición de las fronteras de las capas de la retina que se quieren detectar en este proyecto (ILM;INL-IPL;IS-OS), tras el proceso de entrenamiento y testeo del algoritmo propuesto, se realiza un post-procesado sobre los mapas de segmentación obtenidos, (véase Figura 3.25 (a)). A continuación, se describe cada uno de los pasos por los que está compuesto este proceso.



**Figura 3.25:** Máscaras binarias obtenidas tras cada uno de los pasos de post-procesado: (a) Máscara de segmentación obtenido tras el testeo (b) Máscara binaria tras obtener los componentes con conectividad 8 de cada capa, (c) Máscara binaria tras eliminar los componentes que presentan un área menor al 30% del máximo del área de capa capa (d) Máscara tras realizar un cierre con un disco de radio 30, (e) Mapa de segmentación obtenido.

Como ya se ha comentado en secciones anteriores, para la segmentación de las imágenes de retina, éstas se han dividido en cuatro clases diferentes. Para cada una de las clases de interés, exceptuando las clases del fondo, se obtiene la máscara binaria donde se tendrá a 1 los píxeles detectados por la red como pertenecientes a la capa correspondiente, Figura 3.2.8 (b), (fila superior:primera capa, fila inferior:segunda capa).

A continuación, con el fin de eliminar el posible ruido que no pertenece a la capa retiniana de interés, se calcula para cada clase el número de píxeles totales y se escogen las componentes conexas (conectividad 8) que presenten un número de píxeles igual o superior al 30% del componente con el área máxima para

esa clase. Posteriormente, mientras que el número de objetos o componentes conexas presentes en una clase sea mayor que 1, lo que quiere decir que ha habido una rotura de esa capa, se realiza una operación morfológica denominada cierre con el fin de unirlos. Este tipo de operaciones están fundamentadas en un procesado no lineal donde su elemento base se denomina elemento estructurante,  $B$ , y determina la ventana sobre la que se realiza una operación. La fórmula matemática de la operación de cierre es:

$$\phi_B(x) = \varepsilon_B(\delta_b(X)) \quad (3.13)$$

donde  $\varepsilon_B$  es una operación morfológica conocida como erosión y  $\delta_b$  es otra operación morfológica llamada dilatación. Sus fórmulas matemáticas aparecen en las ecuaciones 3.14 y 3.15 respectivamente:

$$\varepsilon_B(x) = \inf\{X_b, b \in B^*\} = \{z, B_z \in X\} \quad (3.14)$$

$$\delta_B(x) = \sup\{X_b, b \in B^*\} = \{x, B_x \in X\} \quad (3.15)$$

La dilatación es una operación extensiva que, a partir de una imagen binaria y dado uno o más píxeles en blanco, fija con el mismo valor los píxeles del vecindario. El número de píxeles vecinos que se convierten en blanco depende de la conectividad y del elemento estructurante ( $B$ ) elegidos, por tanto, en esta operación el resultado es siempre mayor que la entrada. En cambio, la erosión es una operación morfológica antiextensiva, pues dados una imagen binaria y un conjunto de píxeles en blanco, y siguiendo las mismas propiedades que la dilatación, es capaz de convertir en negro los píxeles más externos de los objetos blancos de la imagen, haciendo que el resultado sea menor que la entrada. Por tanto, el cierre, al calcular primero una dilatación y después una erosión, es una operación extensiva que solo cambia la morfología de las estructuras que son más pequeñas que el elemento estructurante, manteniéndose las demás intactas.

En la aplicación del cierre para conseguir que el número de componentes conexas presentes en una clase sea igual a 1, se comienza con un disco de tamaño 1 como elemento estructurante, y conforme van pasando las iteraciones se va incrementando su tamaño en 1. En cada iteración, esta operación elimina entrantes estrechos y agujeros donde no cabe el elemento estructurante. El resultado de la eliminación de ruido presente en una capa y de la unión de las capas que se encuentren fragmentadas se observa en la Figura (c), aunque cabe destacar que en estas imágenes dado que ninguna de las capas se ha fragmentado, el último paso explicado no se ha llevado a cabo.

Posteriormente, se rellenan los posibles huecos existentes dentro de cada capa y se realiza otro cierre. Esta vez, esta operación se lleva a cabo con un disco como elemento estructurante de tamaño 30 para evitar los entrantes que se generan en los límites de las capas. El efecto de esta operación se puede visualizar en la Figura 3.2.8 (d).

Como resultado del proceso de post-procesado, se obtiene el mapa de segmentación visualizado en la Figura 3.2.8 (e), donde se puede observar que cada una de las capas es definida por completo.

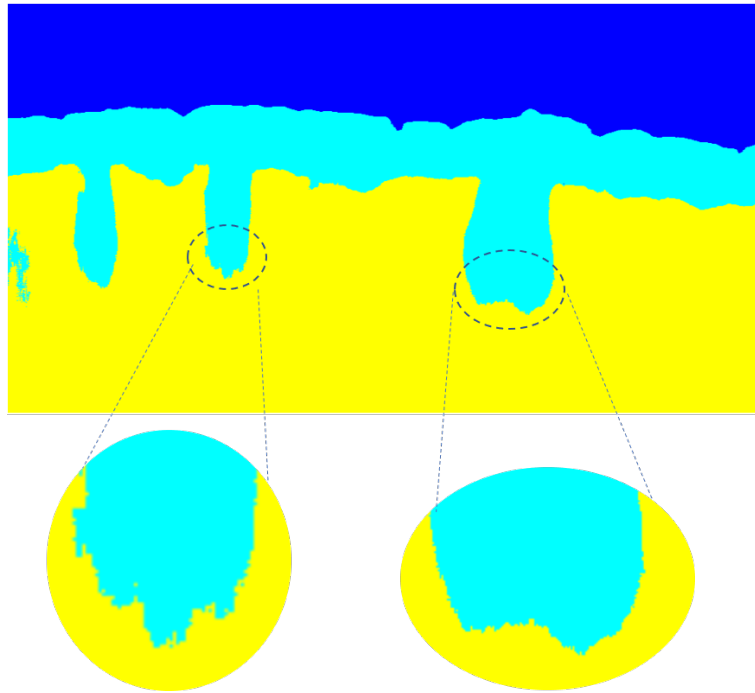
Como se ha comentado, el objetivo del post-procesado es la obtención de cada uno de los límites de algunas de las capas que conforman la retina. Por tanto, posterior al tratamiento de los mapas de segmentación, se obtienen los contornos de cada una de las capas. Debido al ruido *speckle* presente en estas imágenes, resultado del proceso de adquisición, es necesario suavizar las curvas obtenidas de las dife-



rentes capas. Para ello, sobre los puntos pertenecientes a cada capa se ajusta un polinomio de grado 8 y posteriormente se obtiene la curva suavizada con este polinomio.

### *Post-procesado en imágenes epiteliales*

Tras el entrenamiento y testeado del modelo propuesto para la segmentación de las imágenes epiteliales se obtienen los mapas de segmentación visualizados en la Figura 3.26.



**Figura 3.26:** Mapas de segmentación que se obtienen tras el entrenamiento de la red.

Como se puede observar en esta figura, la red no es capaz de definir por completo los bordes pertenecientes a los folículos pilosos debido a la falta de un pronunciado gradiente de intensidad en sus contornos. Para aportar a los folículos su forma característica, se lleva a cabo un post-procesado en los mapas de segmentación obtenidos.

El post-procesado, implementado para mejorar la segmentación que proporciona la red, consta de diferentes partes que pasarán a explicarse a continuación.

#### ■ **Filtrado Savitzky y Golay (SG)**

Como se ha explicado en la sección 2.2, los folículos pilosos son invaginaciones de la epidermis, lo que produce la depresión de esta capa y por tanto, de la unión dermo-epidérmica. En las zonas donde la piel presenta folículos pilosos, la señal correspondiente a esta unión no es de baja frecuencia sino que presenta gran cantidad de variaciones, y por tanto, de mínimos locales. El primer paso para realizar un suavizado de los folículos consiste en la obtención de los contornos de los mismos. Para conseguir este objetivo, es necesario detectar la línea base de la señal dermo-epidérmica. Para ello, se ha utilizado el filtro de Savitzky y Golay.

El filtro Savitzky y Golay, introducido por A. Savitzky y M. J. Golay, en 1964, es considerado un tipo de filtro digital de respuesta finita (FIR)<sup>5</sup> pero más efectivo que estos ya que preserva los componentes de alta frecuencia pertenecientes a la señal [52]. Dicho filtro se basa en el ajuste polinómico de un conjunto de puntos por medio de mínimos cuadrados.

Si se considera una ventana simétrica de tamaño  $N$  sobre un conjunto de datos  $N = 2M + 1$  centrada en el punto  $x_0$ , los puntos contenidos en esta ventana son simétricos a cada lado de  $x_0$  como sigue:

$$x_m = x[x_{-M}, \dots, x_{-1}, x_0, x_1, \dots, x_M]^T \quad (3.16)$$

Las  $N$  muestras contenidas  $x_m$  se pueden ajustar mediante un polinomio de orden  $d$  con ( $0 < d < M$ ), de acuerdo a la siguiente ecuación:

$$\hat{x}_m = c_0 + c_1 m + \dots + c_d m^d \quad -M \leq m \leq M \quad (3.17)$$

donde  $\hat{x}_m$  representa la  $m$ -ésima muestra de datos suavizada. Los coeficientes  $c_i$  son números reales y deben determinarse de manera óptima, minimizando el error del ajuste por mínimos cuadrados:

$$e = \sum_{m=-M}^{m=M} (x_m - \hat{x}_m) = \sum_{m=-M}^{m=M} (x_m - (c_0 + c_1 m + \dots + c_d m^d)) \quad (3.18)$$

Para obtener el valor de estos coeficientes y por tanto de los puntos suavizados, se definen  $d + 1$  vectores base polinomiales representados por  $s_i$  con ( $i = 0, 1, \dots, d$ ) [53]:

$$s_i(m) = m^i \quad -M \leq m \leq M \quad (3.19)$$

La correspondiente matriz  $S$  se establece como:

$$S = [s_0, s_1, \dots, s_d] \quad (3.20)$$

Por tanto, los valores suavizados se pueden organizar en un vector de la siguiente manera:

$$\hat{x} = Bx = \sum_{i=0}^d c_i s_i \quad (3.21)$$

siendo  $B = SG^T$  donde  $G = SF^{-1}$  y  $F = S^t S$  ya que tras el proceso de minimización del error se obtiene que  $c = G^T x$ .

Sin embargo, el valor  $y_0 = \hat{x}_0$  se da en términos del centro del filtro  $b_0$  de acuerdo con la siguiente fórmula:

$$y_0 = Bx = b_0^t x \sum_{m=-M}^M c_0(m) x_i \quad (3.22)$$

---

<sup>5</sup>El filtro FIR consiste en la convolución de la señal de entrada con la respuesta al impulso del filtro. En este caso, la salida es un promedio ponderado de la muestra actual y pasada de la señal de entrada

Además, el vector N-dimensional  $x$  se puede desplazar  $n$  instantes de tiempo de la siguiente forma:

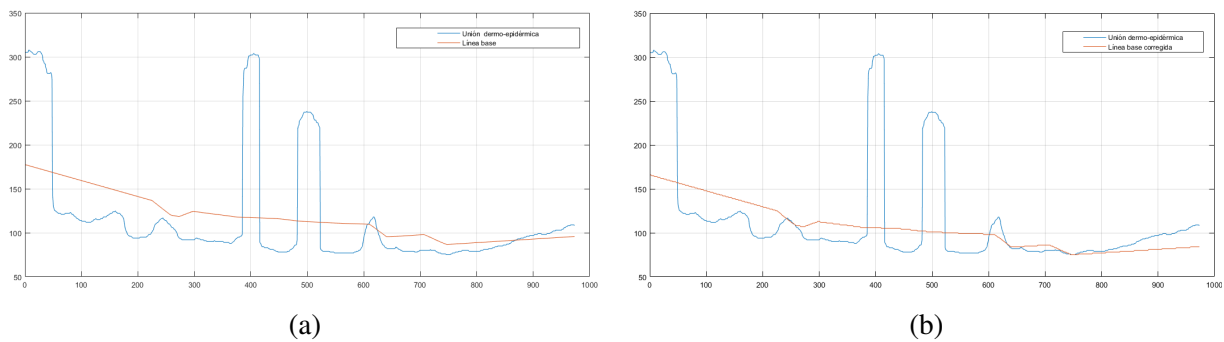
$$x = x[x_{n-M}, \dots, x_{n-1}, x_n, x_{n+1}, \dots, x_{n+M}]^T \quad (3.23)$$

Por lo tanto, la respuesta de un filtro SG con longitud  $N$  y orden  $d$  para suavizar una señal  $x(n)$  puede ser representado por:

$$y(n) = \sum_{m=-M}^M b_0(-m)x(n-m) \quad (3.24)$$

Cuanto mayor sea la longitud de la ventana, menor será la varianza del error de la salida del filtro, pero si esta es demasiado alta, la salida del filtro se sesgará en comparación con la señal real, es decir, la señal de salida se encontrará enormemente suavizada.

Dado que el propósito de la aplicación de este método es la obtención de la línea base de la unión dermo-epidérmica, se aplica un orden de polinomio  $d = 1$  y un tamaño de ventana grande para obtener la señal lo más suavizada posible. Tras la aplicación de este filtro se obtiene la señal roja representada en la imagen (a) de la Figura 3.27.



**Figura 3.27:** Obtención de la imagen de la línea base de la unión dermo-epidérmica: a) Línea base obtenida tras la aplicación del filtro de Savitzky y Golay b) Línea base obtenida tras su corrección.

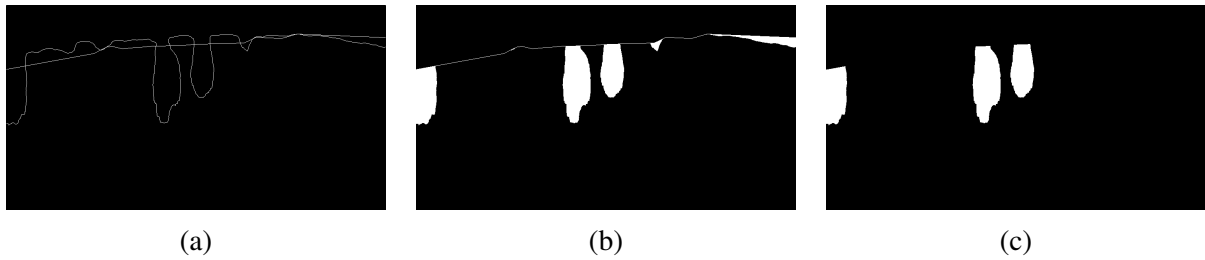
Como se puede observar, la línea base de la unión dermo-epidérmica que se ha obtenido se encuentra desplazada. Por ello, se realiza una corrección de la misma aplicando la siguiente fórmula:

$$lb_{corregida}(x) = lb(x) - dif \quad (3.25)$$

siendo  $dif = \min(signal(x)) - \min(lb(x))$  con  $signal(x)$  la señal de la unión dermo-epidérmica y  $lb$  la línea base obtenida tras aplicar el filtro de Savitzky y Golay.

Una vez corregida la línea base, en la Figura 3.28 (a) se muestra la conversión de las señales pertenecientes a la unión dermo-epidérmica y a la línea base en una imagen.

Como se ha comentado, los folículos son aquellas depresiones pronunciadas de la unión dermo-epidérmica. Por tanto, se consideran como folículos aquellos objetos que se sitúan por debajo de la línea base. De la Figura 3.28 (a), se eliminan todos aquellos componentes que se sitúan por encima de la línea base detectada y se rellenan aquellos contornos cerrados que se sitúan debajo de esta (Véase Figura 3.28 (b)). Sin embargo, como se puede apreciar en esta imagen, hay algunas estructuras que no son folículos pilosos, y



**Figura 3.28:** Obtención de los folículos a partir de la línea base. a) Máscara con la unión dermo-epidérmica y la línea base obtenida b) Máscara con los diferentes objetos situados bajo la línea base c) Imagen con los diferentes folículos.

por tanto, no deben ser consideradas en el análisis de su forma. Para lidiar con este problema, se calcula el área de todos los componentes que presentan conectividad 8 en la Figura 3.28 (b). Los folículos son aquellas estructuras que presentan un área superior al 30% del área del componente con área máxima.

Por tanto, este paso del post-procesado permite diferenciar las estructuras que son folículos para su posterior análisis, Figura 3.28 (c).

#### ■ Descriptores de Fourier

Tras aplicar el filtro de Savitzky y Golay (SG) para obtener la línea base de la señal de la unión dermo-epidérmica y adquirir cada uno de los folículos pilosos, se calculan los descriptores de Fourier de cada uno de ellos con el fin de suavizar su contorno.

Los descriptores de Fourier son una herramienta matemática útil para la descripción de la forma de cualquier objeto definido por una línea de contorno cerrada. El primer paso para la aplicación de la transformada de Fourier que analiza la forma de un folículo piloso, es la obtención de las coordenadas de su contorno  $(x,y)$ , que lo definen con una precisión suficiente. La aplicación del método de Fourier exige además que dicho contorno se discretice en coordenadas equiespaciadas [54]. El número de puntos del contorno debe ser potencia de dos, y su valor limita el número máximo de descriptores que tendrá el objeto. De esta forma, la línea que define el borde del objeto se considera perteneciente al plano complejo, (ver 3.27) y es recorrida a velocidad constante:

$$s[n] = x[n] + jy[n] \quad N \text{ puntos} \quad (3.26)$$

siendo  $x[n]$  e  $y[n]$  las coordenadas de los puntos del contorno remuestreado.

Un vez que se tienen los puntos de contorno en el plano complejo, se calcula la transformada de Fourier sobre ellos.

$$S[k] = \sum_{n=0}^{N-1} s[n] e^{-j \frac{2\pi kn}{N}} \quad (3.27)$$

donde  $k = 0 \dots N - 1$ ,  $f = \frac{k}{N}$  y por tanto,  $f = 0 \dots \frac{N-1}{N}$

Tras el paso citado anteriormente, el contorno del folículo se convierte en un vector de frecuencias. Si se visualiza su espectro se obtiene la Figura 3.29.

La transformada de Fourier tiene inversa, por lo que se puede reconstruir el contorno del objeto a partir de sus descriptores. La reconstrucción se lleva a cabo aplicando la ecuación 3.28:

$$s[n] = \frac{1}{N} \sum_{k=0}^{P-1} s[k] e^{j \frac{2\pi kn}{P}} \quad (3.28)$$

donde  $n = 0 \dots N - 1$

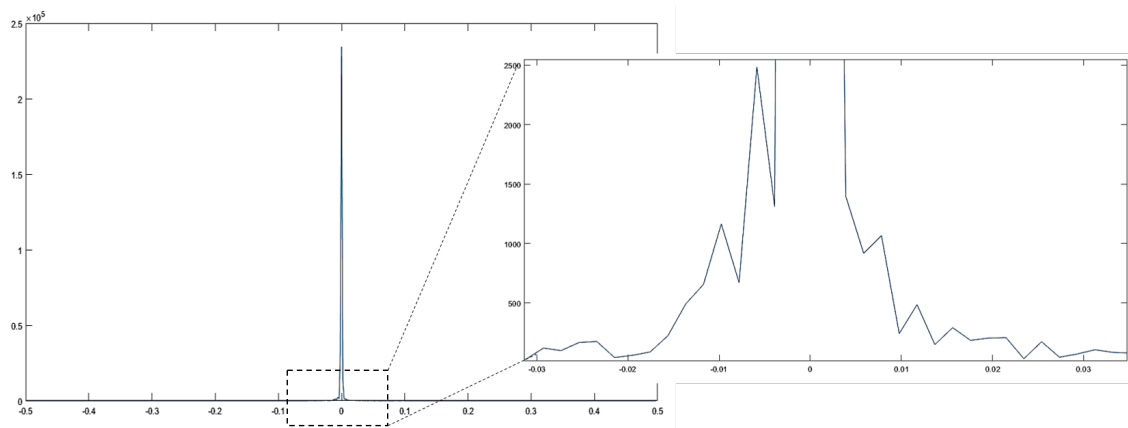


Figura 3.29: Frecuencias contenidas en el folículo piloso sin post-procesado

De todos los descriptores que componen un objeto  $K$ , los primeros, que corresponden a las bajas frecuencias, representan la forma global de su contorno. Los últimos descriptores, que correspondientes a las altas frecuencias, representan los detalles más finos, es decir, la rugosidad digital del contorno. En la Figura 3.30, se ejemplifica a nivel general como influye el número de descriptores considerados  $P$  sobre la información de un contorno.

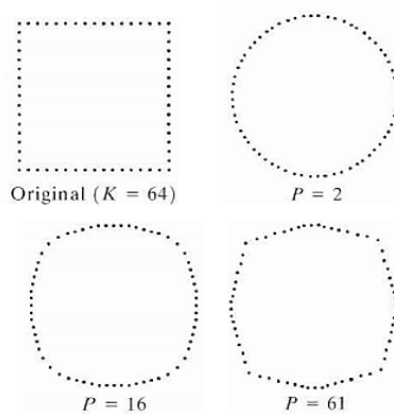


Figura 3.30: Suavizado de un contorno en función del número de descriptores que es escogido

Como se puede observar en la Figura 3.30 a medida que se disminuye el número de descriptores que se tienen en cuenta en el proceso de reconstrucción del objeto, su contorno se suaviza. Sin embargo, a medida que aumenta el número de descriptores el contorno es más similar al que se tenía en el momento de partida. El número total de descriptores a utilizar es  $K$  que es igual a  $N$ , siendo  $N$  el número de puntos totales. La reconstrucción del objeto con este número de descriptores no cambia su forma original.

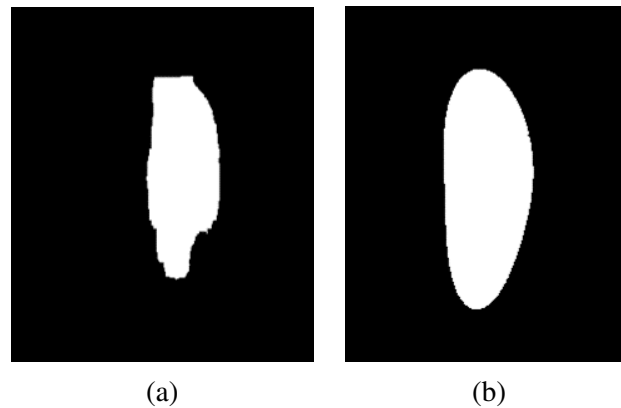
Con respecto al objetivo propuesto, que es el suavizado de la forma que presentan los folículos pilosos obtenidos en el mapa de segmentación, el ancho de banda de la señal folicular es de aproximadamente 0.03 por lo que se utiliza una frecuencia de corte ( $f_c$ ) de aproximadamente 0.007. Esto quiere decir que

todos aquellos descriptores cuyas frecuencias superen  $f_c$  y que por tanto contribuyen a que el contorno presente irregularidades, serán eliminados. Dado que el número de puntos pertenecientes al contorno y por tanto, de descriptores totales en este caso es de  $K = 512$ , el número de descriptores seleccionados puede ser calculado con la siguiente fórmula:

$$f_c = \frac{P}{K} \quad (3.29)$$

En este caso  $P$ , número de descriptores utilizado, es de aproximadamente 4.

En la Figura 3.31 se puede observar cual es el resultado de aplicar este proceso de suavizado sobre un folículo.

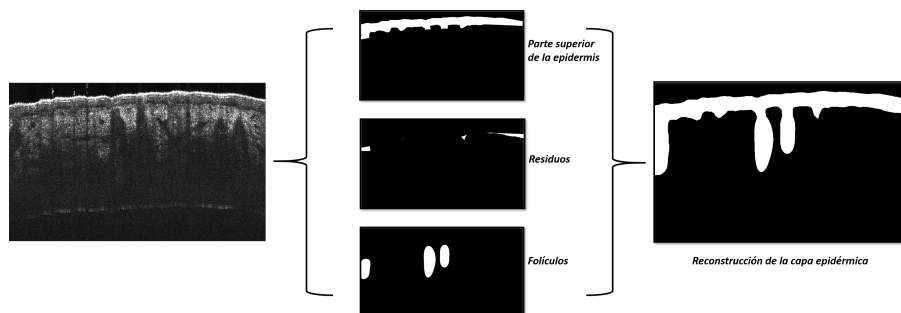


**Figura 3.31:** Post-procesado aplicado para suavizar los folículos: (a) Máscara de un folículo (b) Máscara del folículo suavizado.

Como se puede observar, tras el post-procesado, la forma que tiene el contorno del folículo piloso es más similar a la propuesta en el ground truth de las imágenes de piel.

■ **Reconstrucción de la imagen**

Tras el suavizado del contorno de todos los folículos mediante la aplicación de coeficientes de fourier, es necesario una reconstrucción de la imagen con el fin de obtener la capa epidérmica final. En la Figura 3.32, se ejemplifica este proceso. La máscara final es la suma la capa epidérmica localizada en la parte superior a la línea base, la imagen de residuos (diferencia entre la imagen (b) y (c) perteneciente a la Figura 3.28) y la máscara binaria donde se encuentran los folículos suavizados.



**Figura 3.32:** Reconstrucción de la capa epidérmica a partir de las diferentes imágenes obtenidas tras el proceso de post-procesado en imágenes de piel.

Tras la unión de las máscaras binarias citadas anteriormente, se realiza un cierre para favorecer la cohesión entre las diferentes estructuras.

- **Suavizado de las diferentes capas**

Por último, cabe destacar que al igual que ocurre en las imágenes de retina, estas imágenes se ven afectadas por el ruido denominado *speckle* que puede afectar a los límites de la capa epidérmica de interés. Para suavizar estos límites (partes superior de la epidermis y unión dermo-epidérmica) se diezman ambas señales y posteriormente se interpolan con un polinomio de diferente grado según el límite suavizado.





## Capítulo 4

# Resultados

### *Índice de contenidos*

4.1 Resultados obtenidos en imágenes retinianas . . . . .	61
4.1.1 Resultados cualitativos . . . . .	62
4.1.2 Resultados cuantitativos . . . . .	63
4.1.3 Discusión de los resultados obtenidos en imágenes retinianas . . . . .	64
4.2 Resultados obtenidos en imágenes epiteliales. . . . .	65
4.2.1 Resultados cualitativos . . . . .	65
4.2.2 Resultados cuantitativos . . . . .	66
4.2.3 Discusión de los resultados obtenidos en imágenes epiteliales. . . . .	66

Como se ha comentado a lo largo del proyecto, para comprobar la validez de los resultados obtenidos tras el entrenamiento y testeo de los algoritmos de segmentación automáticos basados en CNN tanto en piel como en retina, se ha realizado una segmentación manual que se propone como *ground truth*. Con esta segmentación manual se han comparado los resultados obtenidos con el fin de determinar el error o el acierto conseguido.

La comparación de los resultados obtenidos con los propuestos en el estado del arte es muy complicada de llevar a cabo debido a diversos factores. Uno de los mayores inconvenientes es la discrepancia entre los tipos de imágenes en las que se aplican los algoritmos para conseguir los objetivos propuestos. Por ejemplo, en el caso de la retina, hay muy pocos estudios en los que se realice la segmentación de las capas en imágenes de roedores sino que la mayoría se realiza sobre humanos. La comparación con estos algoritmos podría presentar discrepancias debido a la diferencia anatómica entre imágenes de roedores y humanos. Con respecto a las imágenes de piel, la mayoría de estudios llevados a cabo son realizados sobre imágenes histopatológicas por lo que la comparación de los resultados obtenidos en imágenes de OCT no sería útil ya que estas modalidades presentan notables diferencias.

Por otro lado, los objetivos de los autores pueden ser diferentes. En el caso de la retina, cuando se compara con el estado del arte, diversos autores desean detectar diferentes estructuras a las detectadas en este proyecto. Por ello, en algunos casos no es posible realizar una comparación directa. En el caso de la piel, tras una larga búsqueda en el estado del arte, no se ha encontrado ningún estudio realizado sobre OCT en el que se incorpore la segmentación de los folículos pilosos debido a que son estructuras difíciles de identificar. Por tanto, si se quisiese comparar con estos estudios el método propuesto en este proyecto podría verse perjudicado aunque esté realizando una tarea más laboriosa.

A continuación, se definen y se explican las medidas que se utilizarán como indicadores para la evaluación de la segmentación automática implementada mediante el uso de redes neuronales convolucionales en las imágenes de capas retinianas.

- **Distancia absoluta entre *boundaries***

Con esta métrica se calcula el error a nivel de curva entre el método que se quiere validar y el *ground truth*. Se expresa mediante la siguiente fórmula:

$$Distancia\ boundaries = \frac{1}{M} \sum_{i=1}^M |gt(i) - y(i)| \quad (4.1)$$

Siendo  $M$  el número de puntos presentes en cada curva,  $gt(i)$  cada uno de los puntos de la curva del *ground truth* e  $y(i)$  los puntos que pertenecen a la curva del método que se quiere validar.

- **Error de grosor**

Esta métrica computa la diferencia de grosor existente entre el *ground truth* y la imagen a evaluar para cada una de las capas segmentadas.

$$Error\ layer = \frac{1}{M} \sum_{i=1}^M |layer_{gt}(i) - layer_{seg}(i)| \quad (4.2)$$

siendo  $M$  el número de píxeles en la dirección  $x$ ,  $layer_{gt}$  el grosor de cada capa del *ground truth* y  $layer_{seg}$  el grosor de cada capa del método que se quiere evaluar.

■ **Dice**

Este índice es un métrica que mide la similitud de una región entre dos imágenes a partir de su intersección [55]. En este caso se computa como el ratio entre la intersección de cada capa segmentada manualmente  $capa_{gth}$  y la segmentada automáticamente por el método a validar  $capa_{seg}$ , y el sumatorio de cada capa.

Su fórmula es la siguiente:

$$S = \frac{2 \sum_{i=1}^{M_i} (capa_{seg} \text{ AND } capa_{gth})}{\sum_{i=1}^{M_i} capa_{seg} + \sum_{i=1}^{M_i} capa_{gth}} \quad (4.3)$$

donde  $M_i$  representa a los píxeles de cada capa. Cabe destacar que  $S$  varía entre 0 y 1.

En este caso, se obtendrán dos valores de *DICE* diferentes puesto que se obtienen tres límites de capas y en su consecuencia dos capas.

■ **Indice de Jacard (JD)**

El índice de Jacard mide la similitud y diversidad de dos muestras a partir de la intersección y la unión de dichas muestras. Matemáticamente se define como la intersección de dos áreas dividida por la unión de las mismas [55].

Su fórmula es la siguiente:

$$JD = \frac{\sum_{i=1}^{M_i} (capa_{seg} \text{ AND } capa_{gth})}{\sum_{i=1}^{M_i} (capa_{seg} \text{ OR } capa_{gth})} \quad (4.4)$$

donde JD toma valores entre 0 y 1.

Cabe destacar que la relación existente entre Dice y Jaccard se puede expresar con la siguiente fórmula:

$$S = 2 * JC(1 + JC) \quad (4.5)$$

Para validar la segmentación de piel se han utilizado los índices de Jaccard y Dice ya que son las métricas más adecuadas para evaluar el tipo de segmentación propuesta en estas imágenes. Las definiciones matemáticas de estas métricas han sido desarrolladas anteriormente.

## 4.1 Resultados obtenidos en imágenes retinianas

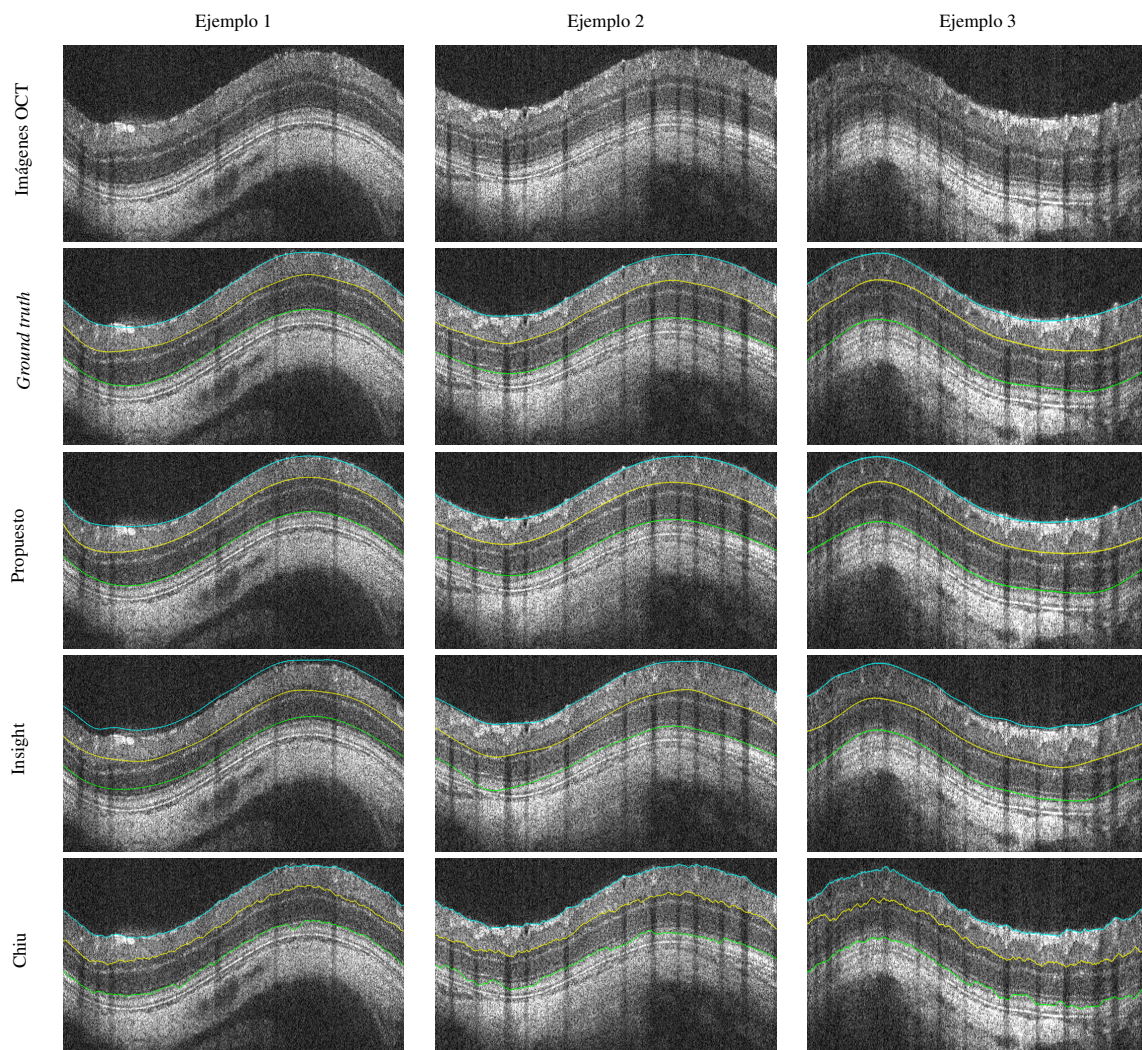
Como se ha expuesto anteriormente, la comparación de los resultados obtenidos con el estado del arte es complicada. Para disminuir en la medida de lo posible la discrepancia en la comparación de resultados, se han testeado las imágenes con las que se ha entrenado y testeado el modelo propuesto en este proyecto con dos métodos diferentes presentes en el estado del arte. Cabe destacar que, para la segmentación de las imágenes retinianas, se contaba con dos entregas diferentes de una base de datos y que los métodos de comparación solo han sido aplicados sobre la primera entrega de la base de datos.

En primer lugar, se ha hecho uso del software *Insight* de Phoenix Research Labs que ofrece una herramienta para la segmentación automática. Cabe destacar que el método *Insight* software no segmenta todas los límites de las capas bajo análisis en todas las imágenes. De las 129 imágenes, el límite de la capa ILM solo es detectado en 113, el límite IPL-INL en 26 y el del segmento IS-OS en 70.

En segundo lugar, se ha utilizado el método de Chiu et al., [56]. Este método está centrado en la segmentación de la OCT humana. Este método ha sido elegido con el objetivo de verificar si un método de código abierto, pues es complicado que los algoritmos estén disponibles, y reproducible que es ampliamente recomendado por la comunidad investigadora para la segmentación de OCT humana es capaz de adaptarse y lograr buenos resultados en las OCT de roedores con las que se trabaja.

#### 4.1.1 Resultados cualitativos

En la figura 4.1 se exponen los resultados obtenidos de forma cualitativa con el fin de obtener una visión general de cómo se han cumplido los objetivos propuestos. En ella, aparece una comparativa de las curvas segmentadas por el método propuesto, el *ground truth* y por los métodos anteriormente mencionados (Insight y Chiu) sobre la primera entrega de la base de datos.



**Figura 4.1:** Los resultados obtenidos de la segmentación se muestran en tres ejemplos representativos de las imágenes OCT pertenecientes a la primera entrega de la base de datos. Código de color: ILM (cian), IPL-INL (amarillo) y IS-OS (verde)

#### 4.1.2 Resultados cuantitativos

Como se ha comentado anteriormente, las métricas utilizadas en este TFG para presentar los resultados cuantitativos son el error de distancia entre boundaries, el error de grosor y el coeficiente de Dice y Jaccard. Cabe destacar que el cálculo de estas métricas se ha efectuado en ( $\mu\text{m}$ ), para ello se ha computado el correspondiente paso de píxeles a ( $\mu\text{m}$ ) multiplicando por la resolución correspondiente. Los resultados del error de distancia entre *boundaries* para las imágenes de la primera y segunda entrega de la base de datos se pueden visualizar en la tablas 4.1 y 4.2 respectivamente.

Boundary	Propuesto	InSight <sup>a</sup>	Chiu et al.
<b>ILM</b>	1,93 ± 0,55	6,97 ± 4,59	2,57 ± 0,58
<b>IPL-INL</b>	3,21 ± 0,72	10,49 ± 4,06	11,50 ± 14,08
<b>IS-OS</b>	2,41 ± 1,11	8,67 ± 5,90	14,75 ± 26,84
<b>Total</b>	<b>2.52 ± 0.8</b>	8,71 ± 4,85	9,61 ± 13,83

**Tabla 4.1:** Errores de distancia absoluta entre curvas (media y desviación estándar en  $\mu\text{ m}$ ) calculados comparando los resultados del método propuesto, InSight y Chiu et al. con el *ground truth* sobre las imágenes de la primera entrega de la base de datos

<sup>a</sup>Cabe destacar que debido al hecho de que el error de distancia absoluta del software InSight fue el mayor entre todos los métodos comparados y no todas las capas están segmentadas, el valor del error y el coeficiente de Dice de este método no se incluye en la tabla 4.3

Boundary	Propuesto
<b>ILM</b>	3,09 ± 0,94
<b>IPL-INL</b>	5,25 ± 3,31
<b>IS-OS</b>	5,04 ± 4,81
<b>Total</b>	4,45 ± 3,02

**Tabla 4.2:** Errores de distancia absoluta entre límites (media y desviación estándar en  $\mu\text{ m}$ ) calculados comparando los resultados del método propuesto con el *ground truth* sobre las imágenes de la segunda entrega de la base de datos

Los resultados de la métrica del error de grosor y del coeficiente de Dice y Jaccard para cada una de las entregas de la base de datos se pueden observar en las tablas 4.3 y 4.4 respectivamente.

Capas	Grosor medio ( $\mu\text{m}$ )			Error de grosor ( $\mu\text{m}$ )		Dice		Jaccard	
	GT	Propuesto	Chiu	Propuesto	Chiu	Propuesto	Chiu	Propuesto	Chiu
<b>RNFL+GCL+IPL</b>	66.81	67.37	73.30	1,7 ± 1,28	8,14 ± 7,95	0,96 ± 0,01	0,93 ± 0,02	0,89 ± 0,098	0,82 ± 0,11
<b>INL+OPL+ONL</b>	97.59	97.44	106.8774	2,14 ± 1,76	11,67 ± 10,75	0,97 ± 0,01	0,94 ± 0,02	0,87 ± 0,11	0,8 ± 0,11
<b>Total</b>	164.41	164.8	180.14	9,9 ± 9,35	1,92 ± 1,52	0,97 ± 0,01	0,94 ± 0,02	0,88 ± 0,1	0,81 ± 0,1

**Tabla 4.3:** Errores de grosor y Dice (media y desviación estándar) comparando los resultados del método propuesto con el *ground truth* (GT) en segunda entrega de la base de datos.

Layer	Grosor medio ( $\mu\text{m}$ )		Error de grosor ( $\mu\text{m}$ )	Dice	Jaccard
	GT	Propuesto	Propuesto	Propuesto	Propuesto
<b>RNFL+GCL+IPL</b>	72.34	67.85	$5,5 \pm 3,8$	$0,94 \pm 0,02$	$0,89 \pm 0,03$
<b>INL+OPL+ONL</b>	95.26	96.89	$4,42 \pm 4,4$	$0,95 \pm 0,03$	$0,9 \pm 0,05$
<b>Total</b>	167.6	164.74	$4,96 \pm 4,1$	$0,9 \pm 0,04$	$0,95 \pm 0,03$

**Tabla 4.4:** Errores de grosor y Dice (media y desviación estándar) comparando los resultados del método propuesto con el *ground truth* sobre la primera entrega de la base de datos.

### 4.1.3 Discusión de los resultados obtenidos en imágenes retinianas

Con respecto a los resultados obtenidos sobre el conjunto de imágenes pertenecientes a la primera entrega de la base de datos, se debe enfatizar que, a diferencia del método propuesto, el software comercial Insight no segmenta un número fijo de límites de capas y además, su segmentación no es lo suficientemente precisa. Es por ello que, en la mayoría de las veces, este método requiere de una segmentación o modificación manual a través de alguna herramienta específica como la que se ha expuesto en la sección 3.2.5.

Los resultados alcanzados por Chiu et al. sobre las imágenes de la primera entrega de la base de datos también son menos precisos en comparación con los obtenidos por el método propuesto. Se debe tener en cuenta que este método no se ha sido diseñado inicialmente para la segmentación de la OCT en roedores, sino para humanos. Por lo tanto, tiene sentido que los mejores resultados se obtengan en la identificación de la ILM ya que en esta capa la diferencia entre roedores y humanos es menos apreciable.

Como se puede apreciar en la sección , el método propuesto en este proyecto para la segmentación de las fronteras de las capas retinianas (ILM; IPL-INL;IS-OS) sobre la primera entrega de la base de datos es el más preciso y el que menor error presenta cuando se compara con el *ground truth*. Este es el motivo por el cual en las tablas 4.2 y 4.4, que muestran los resultados obtenidos sobre la segunda base de datos, solo se ha tenido en cuenta el método propuesto.

Como se ha comentado a lo largo del proyecto, la segunda base de datos retiniana solo se ha utilizado para realizar un segundo testeo del algoritmo y no para fines de entrenamiento. Por tanto, como se ha comentado en secciones anteriores, estas imágenes son completamente nuevas para el modelo generado por el algoritmo. Este experimento simula las condiciones que se tendrían en la práctica clínica, donde el algoritmo tendría que predecir nuevas imágenes con características disimilares a aquellas con las que ha sido entrenado.

Aunque los resultados obtenidos por el método propuesto sobre la segunda entrega de la base de datos no son tan buenos como los obtenidos sobre la primera entrega, siguen siendo prometedores y demuestran que los modelos generados por el algoritmo desarrollado en este proyecto pueden generalizarse y aplicarse a diferentes conjuntos de datos. Una de las causas a las que podría deberse este mayor error obtenido es que las imágenes pertenecientes a la segunda entrega de la base de datos presentan gran variabilidad y en algunos casos, artefactos debidos a una incorrecta adquisición de la imagen que pueden afectar al buen funcionamiento de la red, (véase Figura 4.2).

Como se ha comentado anteriormente, las mayores desventajas del método que se propone en este proyecto, basado en redes neuronales convolucionales, es la necesidad de realizar una segmentación manual de las capas de interés así como el potente hardware necesario para poner entrenar una red del que normalmente no se dispone en el ámbito clínico. Sin embargo, como se ha demostrado durante el

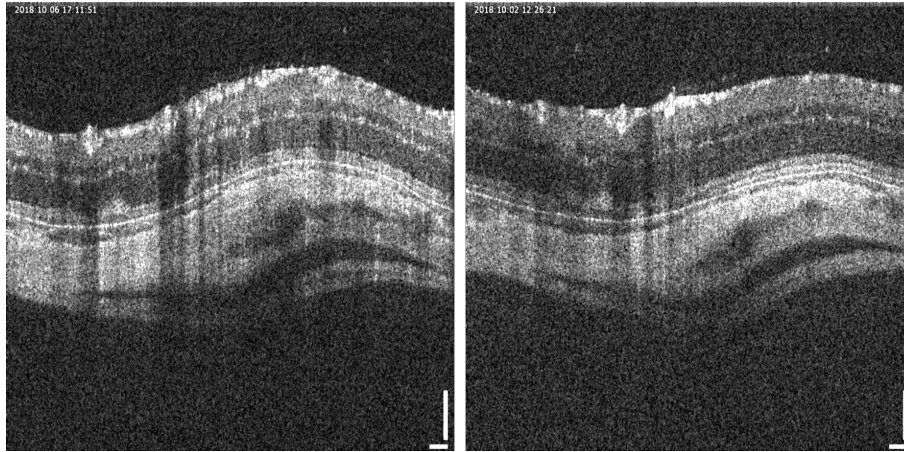


Figura 4.2: Imágenes representativas de la segunda entrega de la base de datos

segundo testeo del modelo propuesto con la segunda entrega de la base de datos, si previamente se ha entrenado el algoritmo con una gran variedad de imágenes es posible realizar una segmentación sin la necesidad de gran cantidad de tiempo ni de un hardware específico.

## 4.2 Resultados obtenidos en imágenes epiteliales

### 4.2.1 Resultados cualitativos

En la figura 4.3 se pueden comparar a nivel cualitativo los resultados del método propuesto frente a la segmentación del *ground truth* para las imágenes de piel.

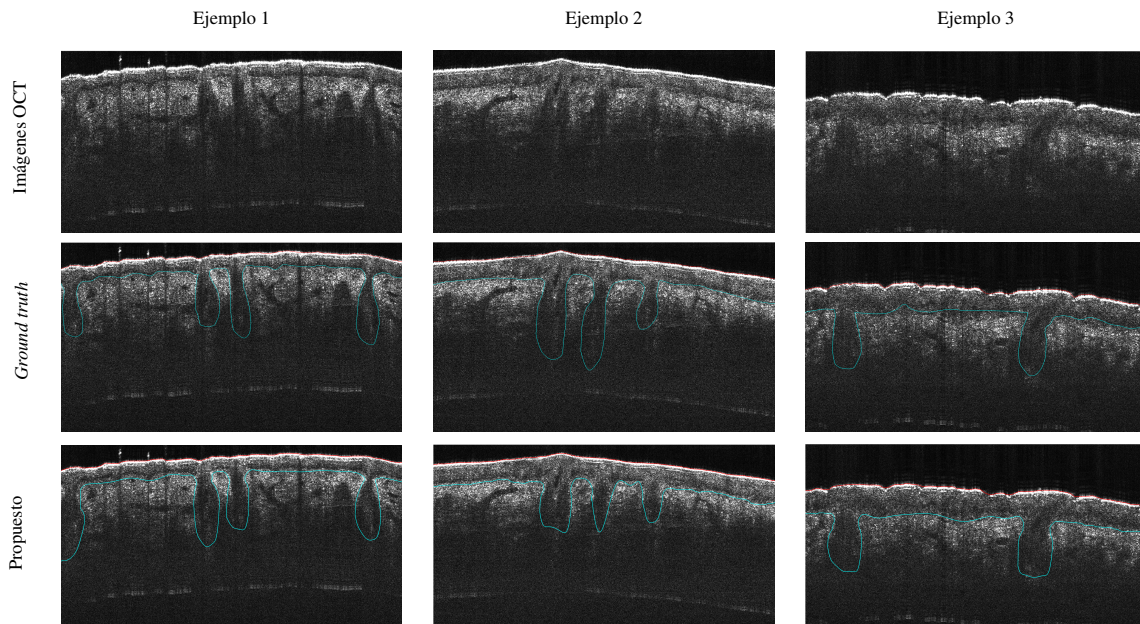


Figura 4.3: Los resultados obtenidos de la segmentación de la epidermis se muestran en tres ejemplos representativos de las imágenes OCT. Código de color: superficie de la piel (rojo), Unión unión dermo-epidérmica (azul)

#### 4.2.2 Resultados cuantitativos

Para analizar los resultados obtenidos sobre las imágenes de piel de forma cuantitativa se han computado las métricas de Dice y el coeficiente de Jaccard sobre la capa de la epidermis y de la dermis. Estos resultados pueden visualizarse en la Tabla 4.5.

Layer	Dice	Jaccard
Epidermis	$0,83 \pm 0,062$	$0,71 \pm 0,13$
Dermis	$0,96 \pm 0,014$	$0,86 \pm 0,16$

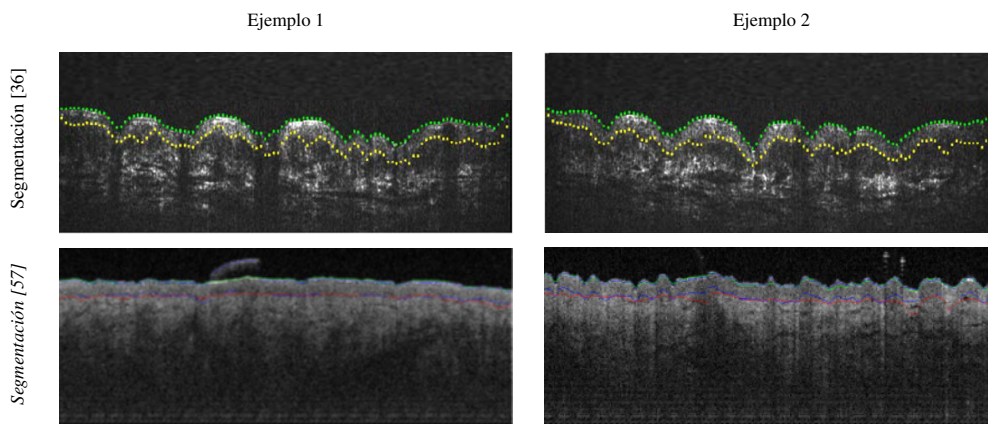
**Tabla 4.5:** Índice Dice y Jaccard (media y desviación estándar) comparando los resultados del método propuesto con el *ground truth*.

#### 4.2.3 Discusión de los resultados obtenidos en imágenes epiteliales

Como se ha explicado en la sección 2.2, los folículos pilosos son invaginaciones del tejido epidérmico, concretamente surgen por el decaimiento de la unión dermo-epidérmica. Se considera que la epidermis va por debajo de los folículos pilosos y por ello, estos se ha considerado pertenecientes a la región epidérmica.

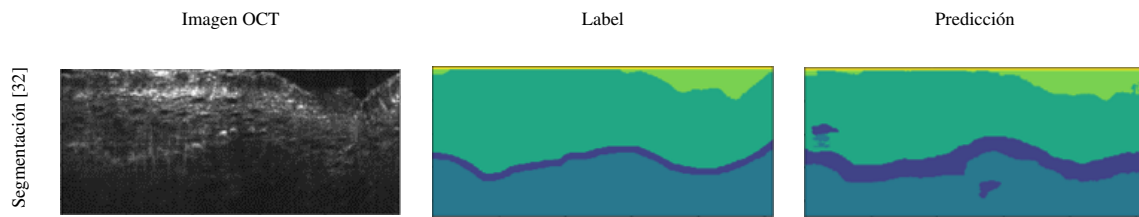
Como se puede observar en la Tabla 4.5 se han calculado las métricas propuestas tanto para la epidermis como para la dermis, puesto que todo el tejido que se observa en la figura 4.3, debajo de la unión dermo-epidérmica pertenece a la dermis. Como se ha explicado en la sección 2.2, debajo de la dermis hay otra capa, la capa subcutánea, pero esta es demasiado profunda para ser visualizada con OCT, por lo que no puede detectarse. La razón por la que las métricas computadas tienen un valor menor en la epidermis que en la dermis se debe en que en algunos casos, el contorno del folículo piloso es difuso y no puede ser detectado con elevada precisión.

Los resultados obtenidos no pueden ser extrapolados y comparados de manera fidedigna con los existentes en el estado del arte ya que es la primera vez que en la modalidad OCT se lleva a cabo el reconocimiento de los folículos pilosos. Como se puede observar en las figuras 4.4 y 4.5 existen algunos estudios, basados en técnicas clásicas, solo uno basado en *deep learning*, en los que se lleva a cabo la segmentación de la capa epidérmica. Sin embargo, en estas imágenes existen folículos pilosos que no se han tenido en cuenta al realizar la segmentación.



**Figura 4.4:** Visualización cualitativa de la segmentación de la epidermis basada en técnicas clásicas llevada a cabo por [36] y [57].





**Figura 4.5:** Visualización cualitativa de la segmentación de la epidermis basada en redes neuronales convolucionales llevada a cabo por [37]. Fondo (verde claro), epidermis (verde oscuro), unión dermo-epidérmica (azul oscuro), dermis (azul).

El cierto que en imágenes de OCT es complicado llevar a cabo un análisis de los folículos pilosos debido a que en algunos casos no existe un claro gradiente que identifique su contorno. Sin embargo, existen algunas enfermedades tales como la hidrosadenitis, la alopecia o el acné en las que el análisis de las estructuras foliculares resulta fundamental para su detección y seguimiento. Es por ello que parte de este proyecto se ha focalizado en la detección de los folículos pilosos incorporando para ello un potente post-procesado.



## Capítulo 5

# Conclusiones y líneas futuras

### *Índice de contenidos*

5.1 Conclusiones . . . . .	70
5.2 Líneas futuras . . . . .	71

## 5.1 Conclusiones

La principal motivación de este proyecto reside en la difusión de nuevas técnicas de diagnóstico por imagen que suplan todos los inconvenientes que presentan las técnicas que se utilizan en la actualidad. Una de las tecnologías que presenta un enorme desarrollo por todas sus ventajas en el análisis de estructuras internas es la OCT. Es por ello que, en este trabajo se han abordado técnicas de procesado, como es la segmentación, para que se apliquen a las imágenes obtenidas con este dispositivo.

Para llevar a cabo la segmentación en imágenes de OCT, primeramente se ha realizado una revisión en la literatura de las principales técnicas que se venían utilizando en este tipo de imágenes, así como cuáles son las que ocupan el estado del arte actual tanto en retina como en piel, puesto que son los dos órganos de estudio abordados en este trabajo. Tras realizar una amplia búsqueda, se ha comprobado que el trabajo realizado en estos dos ámbitos no es amplio.

Previo al diseño de una arquitectura de red neuronal para conseguir el objetivo propuesto, se han introducido los conceptos fundamentales necesarios para entender las partes por las que pueden estar formadas estos tipos de algoritmos. Posteriormente, se ha propuesto la arquitectura adecuada para la segmentación de las estructuras de interés, destacando que debido a las diferentes características que presentan las imágenes de piel y retina las arquitecturas encoder-decoder a implementar han sido ligeramente diferentes.

El algoritmo elegido para llevar a cabo el proceso de segmentación es supervisado, por lo que requiere de la presencia de las etiquetas de las imágenes, es decir, de una segmentación manual de las mismas. Dado que no se contaba con una segmentación previa o si se contaba debía ser modificada, para el desarrollo del proyecto se ha implementado una aplicación de segmentación manual en el entorno de Matlab. Con ella, se ha modificado la segmentación realizada en retina y se ha desarrollado por completo y, con la supervisión de personal especializado, la segmentación en piel.

Una vez que se ha contado con las imágenes y su correspondiente etiqueta, se ha implementado una técnica de partición de datos, denominada *K-fold cross validation*, de manera externa e interna con  $K=8$  para las imágenes de retina y con  $k=9$  para las de piel. Para ello, se han dividido cada una de las bases de datos con las que se ha contado en subconjuntos que aproximadamente presentan el mismo número de imágenes. En cada iteración  $K$ , un subconjunto se ha utilizado para test y los restantes para la fase de entrenamiento. Dentro de la fase de entrenamiento en cada iteración  $K - 1$  se ha empleado un conjunto distinto para validación y los restantes para el entrenamiento propiamente dicho.

Tras la partición de los datos se ha llevado a cabo el proceso de entrenamiento. Para el entrenamiento, tanto en imágenes de retina como en imágenes de piel, se ha utilizado una función de pérdidas logística ponderada de múltiples clases, la *cross-entropy ponderada*. La minimización de esta función se ha llevado a cabo mediante el algoritmo de descenso por gradiente estocástico con la modificación de *momentum*.

Tras el testeo y la obtención de los mapas de segmentación con las dos bases de datos, se ha llevado a cabo un proceso de post-procesado. Cabe destacar que este ha sido diferente en imágenes de piel y retina ya que los objetivos que tenía que suplir este proceso eran notablemente diferentes. Mientras que en imágenes de retina se pretendía mejorar la definición de los *boundaries* de las diferentes capas, en imágenes de piel se pretendía aportar la forma característica de los folículos pilosos en el caso que no la tuvieran.

Posteriormente a la aplicación del post-procesado se han obtenido diferentes resultados tanto cualitativos como cuantitativos utilizando diferentes métricas en el caso de los cuantitativos. Además, para el caso de imágenes de retina se ha intentado hacer una comparación con algunos de los métodos presentes en el estado del arte. En el caso de imágenes de piel este hecho no ha sido posible debido a que tras una búsqueda en el estado del arte no se ha encontrado ningún trabajo que incorpore en su segmentación los folículos pilosos en imágenes de OCT.

Finalmente, puesto que se ha contado con una segunda entrega de la base de datos de retina, se ha pretendido realizar una simulación de la práctica clínica real en la que llegan nuevas imágenes que el algoritmo nunca antes había visto y sin necesidad de reentrenar el mismo, se realiza una predicción de las nuevas imágenes. Los resultados conseguidos sobre estas imágenes son realmente prometedores y demuestran que el método desarrollado puede ser generalizado.

En definitiva, se puede concluir que se ha logrado diseñar y desarrollar un sistema automático de segmentación en imágenes de retina y piel con el objetivo de detectar aquellas estructuras que son determinantes para el diagnóstico precoz de determinadas enfermedades.

Cabe destacar que como resultado de este proyecto se ha desarrollado el paper: “Towards Automatic Glaucoma Assessment: An Encoder-decoder CNN for Retinal Layer Segmentation in Rodent OCT images” que ha sido aceptado en el congreso internacional de procesamiento de imagen (EUSIPCO) 2019.

## 5.2 Líneas futuras

Como se ha comentado a lo largo de este proyecto, este trabajo se enmarca dentro del proyecto europeo GALAHAD por lo que las líneas futuras de este trabajo se ligan a sus objetivos finales. GALAHAD promueve el desarrollo de un prototipo de tomografía de coherencia óptica con más resolución y con un menor coste que los que existen actualmente. Esto permitiría una mayor extensión de esta técnica para el diagnóstico precoz de muchas enfermedades como las que se citan en este proyecto.

Con respecto a las imágenes de retina, la línea de investigación debería seguir el objetivo principal del proyecto GALAHAD que es la detección del glaucoma. Tras el desarrollo de este trabajo, se ha conseguido el primer paso para llevarlo a cabo, la segmentación de las regiones de interés que permiten realizar un diagnóstico precoz de esta enfermedad. Por tanto, en trabajos posteriores debería contemplarse el desarrollo de modelos de clasificación supervisada que permitiesen detectar si una imagen presenta o no Glaucoma.

El objetivo de la clasificación consiste en la asignación de un objeto de las diversas categorías o clases especificadas. En concreto, en la clasificación supervisada los algoritmos son entrenados con características de datos que presentan una clase concreta la cual previamente se conoce. De esta forma, los modelos de clasificación pueden ser capaces de predecir la categoría a la que pertenecen nuevas muestras a partir del cálculo de las características seleccionadas durante el entrenamiento. En concreto, en este caso la principal característica para entrenar el modelo sería el grosor de la capas (RNLF+GCL+IPL) el cual se ha obtenido mediante algoritmos de segmentación automática. Como se ha comprobado a lo largo del proyecto, el grosor de estas capas es la principal característica que permite realizar un diagnóstico precoz de la enfermedad. También se podría introducir al modelo otras características que pueden llegar a ser relevantes en esta enfermedad tales como la edad, el campo visual etc. Las diferentes clases que se querrían identificar con este algoritmo serían la ausencia de glaucoma, una fase temprana de la enfermedad o una fase avanzada. Por tanto, se querría abordar un problema de clasificación multiclase.

Para el desarrollo de este algoritmo se haría una revisión bibliográfica de las técnicas que mejores resultados presentan en este tipo de problemas y se probaría con diferentes modelos hasta obtener un resultado prometedor. Tras esto, con el fin de introducir el trabajo desarrollado en la práctica clínica, se desarrollaría una aplicación que permita cargar diferentes imágenes de retina, realizar una segmentación automática con el algoritmo desarrollado en este proyecto, calcular su grosor y por último, predecir si se trata de una imagen que presenta glaucoma o no.

Con respecto a las imágenes de piel, como se ha comentado a lo largo del proyecto, en primer lugar, sería interesante realizar un segundo testeo con imágenes pertenecientes a diferentes pacientes para validar el comportamiento de la red propuesta en este proyecto. Por ello, es necesario una continua comunicación entre los diferentes *partners* que participan en el proyecto Galahad.

Por otro lado, la base de datos de piel con la que se ha contado, solo contiene imágenes de pacientes sanos. La previa segmentación de este tipo de imágenes es importante para el diagnóstico de enfermedades epiteliales. La extracción del grosor de la capa epidérmica a partir de la segmentación de pacientes sanos es importante ya que constituye el desarrollo de un biomarcador que actúa como referencia. Como línea futura en este tipo de imágenes, tendría cabida el entrenamiento del algoritmo desarrollado con imágenes patológicas para poder determinar el comportamiento del algoritmo sobre este tipo de imágenes. En el caso de que el algoritmo desarrollado en este proyecto no actúe bien sobre este tipo de imágenes, se cambiará la configuración de la arquitectura hasta obtener los resultados deseados. Esto permitirá comparar los diferentes grosores entre imágenes sanas y patológicas y establecer las diferencias entre este tipo de imágenes.

# Bibliografía

- [1] OMS, <https://www.who.int/es/news-room/fact-sheets/detail/blindness-and-visual-impairment>, 2018.
- [2] El Mundo, <https://www.cancer.org/es/cancer/cancer-de-piel-de-celulas-basales-y-escamosas/acerca/estadisticas-clave.html>, 2019.
- [3] ———, <https://www.elmundo.es/salud/2015/04/15/552e7aea268e3efe5f8b456e.html>, 2015.
- [4] International Federation of psoriasis associations, <https://ifpa-pso.com/our-actions/world-psoriasis-day>.
- [5] U. Baran, W. J. Choi y R. K. Wang, “Potential use of OCT-based microangiography in clinical dermatology”, *Skin research and technology*, vol. 22, n.º 2, págs. 238-246, 2016.
- [6] E. S. Park, “Skin-layer analysis using optical coherence tomography (OCT)”, *Medical Lasers; Engineering, Basic Research, and Clinical Application*, vol. 3, n.º 1, págs. 1-4, 2014.
- [7] B. Pharma, <https://www.brillpharma.com/anatomia/partes-del-ojo/>.
- [8] A. Piñero, *Aparato Ocular*. Pharma Consult, 1992.
- [9] A. Colomer, *Fundus image analysis for automatic screening of ophthalmic pathologies*. Universidad Politécnica de Valencia, 2018.
- [10] F. Armadá Maresca, A Fonseca Sandomingo, J. Encinas Marn, J García Arumí, F. Gómez-Ulla de Irazazábal, J. Ruiz Moreno y D Vilaplana, “Patología y cirugía de la mácula”, *Sociedad española de Oftalmología*, págs. 27-161, 2010.
- [11] W. L. M. Alward, *Glaucoma: los requisitos en oftalmología*. Harcourt-Mosby, 2000.
- [12] G. Julián, F. Honrubia y L. P. Júlvez, *Diagnóstico y tratamiento del glaucoma de ángulo abierto*. Sociedad Española de Oftalmología, 2003.

- [13] J. Reiriz, *TEJIDOS. MEMBRANAS. PIEL. DERIVADOS DE LA PIEL*. Col·legi Oficial d'Infermeres i Infermers de Barcelona, 2004.
- [14] G. N. Franco, "Histología de la piel", *UNAM*, vol. 46, n.º 4, págs. 130-3, 2003.
- [15] A. Phillips, I. Teo y J. Lang, "Segmentation of Prognostic Tissue Structures in Cutaneous Melanoma Using Whole Slide Images", en *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019.
- [16] J. Krueger y A Bowcock, "Psoriasis pathophysiology: current concepts of pathogenesis", *Annals of the rheumatic diseases*, vol. 64, n.º suppl 2, págs. 30-36, 2005.
- [17] E. Chamorro, "Fundamentos ópticos de la Tomografía de Coherencia Óptica (OCT)", *Ver y Oír*, vol. 231, ene. de 2008.
- [18] J. A. Izatt, M. A. Choma y A.-H. Dhalla, "Theory of optical coherence tomography", *Optical Coherence Tomography: Technology and Applications*, págs. 65-94, 2015.
- [19] A. Boixadera Espax, *Descripción y análisis de los hallazgos con tomografía de coherencia óptica en el desprendimiento de la retina rregmatógeno primario con afectación macular*. Universitat Autònoma de Barcelona, 2004.
- [20] J. M. S. González, "Tomografía de coherencia óptica: técnicas avanzadas en aplicaciones clínicas de la fisiología ocular", *Gaceta de optometría y óptica oftálmica*, n.º 500, págs. 64-71, 2015.
- [21] D. López Quiroga, "Realización de un Tomógrafo Óptico Coherente", 2007.
- [22] C. Griño, F. L. Quintas, M. León, J. M. R. Moreno, S. Ligeró y J. M. Moreno, "Tomografía de Coherencia Óptica (OCT): Funcionamiento y utilidad en patología macular (I)", *Gaceta óptica: Órgano Oficial del Colegio Nacional de Ópticos-Optometristas de España*, n.º 427, págs. 12-14, 2008.
- [23] N. M. Israelsen, M. Maria, M. Mogensen, S. Bojesen, M. Jensen, M. Haedersdal, A. Podoleanu y O. Bang, "The value of ultrahigh resolution OCT in dermatology - delineating the dermo-epidermal junction, capillaries in the dermal papillae and vellus hairs", *Biomed. Opt. Express*, vol. 9, n.º 5, págs. 2240-2265, 2018.
- [24] M. N. Moreno, *Base de datos normativa del grosor de las capas internas de la retina medido con OCT-Spectralis*. Universidad Complutense de Madrid, 2018.
- [25] J. Weissman, T. Hancewicz y P. Kaplan, "Optical coherence tomography of skin for measurement of epidermal thickness by shapelet-based image analysis", *Opt. Express*, vol. 12, n.º 23, págs. 5760-5769, 2004. DOI: 10.1364/OPEX.12.005760.



- [26] J. Welzel, “Optical coherence tomography in dermatology: a review”, *Skin Research and Technology: Review article*, vol. 7, n.º 1, págs. 1-9, 2001.
- [27] MathWorks, [https://es.mathworks.com/help/matlab/learn\\_matlab/product-description.html](https://es.mathworks.com/help/matlab/learn_matlab/product-description.html).
- [28] MatConvNet, <http://www.vlfeat.org/matconvnet/>.
- [29] NVIDIA cuDNN, <https://developer.nvidia.com/cudnn>.
- [30] MobaXterm, <https://mobaxterm.mobatek.net/>.
- [31] Nvidia, <https://developer.nvidia.com/discover/convolutional-neural-network>.
- [32] J. Haggerty, X.-n. Wang, A. Dickinson, C. O’Malley y E. B Martin, “Segmentation of epidermal tissue with histopathological damage in images of haematoxylin and eosin stained human skin”, *BMC medical imaging*, vol. 14, pág. 7, feb. de 2014. DOI: 10.1186/1471-2342-14-7.
- [33] A. Yazdanpanah, G. Hamarneh, B. R. Smith y M. V. Sarunic, “Segmentation of Intra-Retinal Layers From Optical Coherence Tomography Images Using an Active Contour Approach”, *IEEE Transactions on Medical Imaging*, vol. 30, n.º 2, págs. 484-496, 2011, ISSN: 0278-0062. DOI: 10.1109/TMI.2010.2087390.
- [34] P. P. Srinivasan, S. J. Heflin, J. A. Izatt, V. Y. Arshavsky y S. Farsiu, “Automatic segmentation of up to ten layer boundaries in SD-OCT images of the mouse retina with and without missing layers due to pathology”, *Biomed. Opt. Express*, vol. 5, n.º 2, págs. 348-365, 2014. DOI: 10.1364/BOE.5.000348.
- [35] L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li y S. Farsiu, “Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search”, *Biomedical optics express*, vol. 8, n.º 5, págs. 2732-2744, 2017.
- [36] A. Li, J. Cheng, A. P. Yow, C. Wall, D. W. K. Wong, H. L. Tey y J. Liu, “Epidermal segmentation in high-definition optical coherence tomography”, en *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, págs. 3045-3048. DOI: 10.1109/EMBC.2015.7319034.
- [37] M. Calderon-Delgado, J. Tjiu, M. Lin y S. Huang, “High Resolution Human Skin Image Segmentation by means of Fully Convolutional Neural Networks”, en *2018 International Conference on Numerical Simulation of Optoelectronic Devices (NUSOD)*, 2018, págs. 31-32. DOI: 10.1109/NUSOD.2018.8570241.
- [38] J. Brownle, <https://machinelearningmastery.com/batch-normalization-for-training-of-deep-neural-networks>, 2019.

- [39] H. Jansma, <https://towardsdatascience.com/intuit-and-implement-batch-normalization-c05480333c5b>, 2018.
- [40] A. G. Roy, S. Conjeti, D. Sheet, A. Katouzian, N. Navab y C. Wachinger, “Error Corrective Boosting for Learning Fully Convolutional Networks with Limited Data”, en *Medical Image Computing and Computer Assisted Intervention MICCAI 2017*, M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins y S. Duchesne, eds., Springer International Publishing, 2017, págs. 231-239, ISBN: 978-3-319-66179-7.
- [41] V. Badrinarayanan, A. Kendall y R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, n.º 12, págs. 2481-2495, 2017, ISSN: 0162-8828. DOI: 10.1109/TPAMI.2016.2644615.
- [42] A. G. Roy, S. Conjeti, D. Sheet, A. Katouzian, N. Navab y C. Wachinger, “Error Corrective Boosting for Learning Fully Convolutional Networks with Limited Data”, en *Medical Image Computing and Computer-Assisted Intervention MICCAI 2017*, Cham: Springer International Publishing, 2017, págs. 231-239, ISBN: 978-3-319-66179-7.
- [43] F. Sancho, *Entrenamiento de Redes Neuronales: mejorando el Gradiente Descendiente*, 18 de Junio de 2018, 2017.
- [44] A. Ng, <http://deeplearning.stanford.edu/tutorial/supervised/OptimizationStochasticGradientDescent/>.
- [45] M. Sanho, *Diseño de Comportamientos Robóticos Mediante deep Learning*. 2017.
- [46] A. G. Roy, S. Conjeti, S. P. K. Karri, D. Sheet, A. Katouzian, C. Wachinger y N. Navab, “Re-LayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks”, *Biomed. Opt. Express*, vol. 8, n.º 8, págs. 3627-3642, 2017. DOI: 10.1364/B0E.8.003627.
- [47] O. Ronneberger, P. Fischer y T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”, en *Lecture Notes in Computer Science*, Springer International Publishing, 2015, págs. 234-241. DOI: 10.1007/978-3-319-24574-4\_28.
- [48] S. Chilamkurthy, <http://blog.qure.ai/notes/semantic-segmentation-deep-learning-review>, 2018.
- [49] S. Ioffe y C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift”, en *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, Lille, France, 2015, págs. 448-456.
- [50] MatchWorks, <https://es.mathworks.com/discovery/supervised-learning.html>.
- [51] Gracia Quesada, Jesús, *Interpolación: Splines cúbicos*. 2000.

- [52] W Gander, *Solving Problems in Scientific Computing Using Maple and MATLAB*. Springer-Verlag Berlin Heidelberg, 2004.
- [53] M. A. De Oliveira, N. V. S. Araujo, R. N. Da Silva, T. I. Da Silva y J. Epaarachchi, “Use of Savitzky–Golay Filter for Performances Improvement of SHM Systems Based on Neural Networks and Distributed PZT Sensors”, *Sensors*, vol. 18, n.º 1, 2018. DOI: 10.3390/s18010152.
- [54] A. A.S. L. González E. Feliú V., “Descriptores de fourier para la identificación y posicionamiento de contornos en 3D”, 2004.
- [55] H.-H. Chang, A. Zhuang, D. Valentino y W.-C. Chu, “Performance Measure Characterization for Evaluating Neuroimage Segmentation Algorithms”, *NeuroImage*, vol. 47, págs. 122-35, mayo de 2009. DOI: 10.1016/j.neuroimage.2009.03.068.
- [56] S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt y S. Farsiu, “Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation”, *Opt. Express*, vol. 18, n.º 18, págs. 19 413-19 428, 2010. DOI: 10.1364/OE.18.019413.
- [57] H. Urban y T. Sixta, “Epidermis/dermis boundary segmentation in OCT skin images”, 2010.



**Parte II**

**Presupuesto.**



## Capítulo 1

# Presupuesto

### *Índice de contenidos*

1.1 Objetivo . . . . .	82
1.2 Presupuestos parciales . . . . .	82
1.2.1 Costes de personal . . . . .	82
1.2.2 Costes de hardware . . . . .	82
1.2.3 Costes de software . . . . .	84
1.3 Presupuesto Total . . . . .	84

## 1.1 Objetivo

El objetivo de esta sección es aportar una valoración económica de este proyecto basado en el desarrollo de sistemas de segmentación automáticos aplicados a imágenes OCT de retina y piel.

## 1.2 Presupuestos parciales

Con el objetivo de cuantificar los presupuestos parciales, estos se han desglosado en: (i) coste de personal, (ii) coste de *software* y (iii) coste de *hardware*. Este desglosamiento ofrece la posibilidad de realizar de forma rápida un recalcdo del coste total si alguna condición varía.

### 1.2.1 Costes de personal

En este apartado se tienen en cuenta los recursos humanos que han sido necesarios para el correcto desarrollo del proyecto.

En la Tabla 1.1 se desglosan los costes de mano de obra, mostrando la remuneración de cada uno de los participantes en base al tiempo dedicado al proyecto. Concretamente, en este proyecto han participado:

- *D<sup>a</sup> Valery Naranjo Ornedo*, catedrática de universidad y tutora del presente trabajo.
- *D. Adrián Colomer Granero*, doctor de la universidad y encargado de cotutorizar el proyecto.
- *D<sup>a</sup> Rocío del Amor del Amor*, estudiante del Grado de Ingeniería Biomédica y autora del proyecto.

Descripción	Uds.	Cantidad	Precio unitario (€/h)	Coste imputable (€)
Catedrática de universidad	h	32	42	1.344,00
Doctor	h	32	20,50	656,00
Estudiante	h	300	12,50	3.750,00
<b>TOTAL</b>				<b>5.750,00 €</b>

**Tabla 1.1:** Descomposición de los costes de personal.

### 1.2.2 Costes de hardware

Como se ha explicado en secciones anteriores, la mayor parte del proyecto se ha desarrollado utilizando un único ordenador. Sin embargo, dado que el entrenamiento de redes neuronales convolucionales requiere una potencia computacional elevada, este se ha llevado a cabo en uno de los servidores de computación del grupo de investigación CVB Lab. Se han entrenado dos arquitecturas diferentes, una para las imágenes de piel y otra para las de retina. El entrenamiento de cada una de estas redes dura aproximadamente entre cuatro y cinco días. Dado que previamente se realizaron algunas pruebas con diferentes topologías, se puede concluir que se ha hecho uso de este servidor aproximadamente un mes. Además, durante este tiempo se ha hecho uso de un servidor de datos para el almacenamiento de todas las imágenes utilizadas.

En la Tabla 1.2 se calcula el coste que ha supuesto el uso del equipo personal, mientras que en la Tabla 1.3 se muestra el coste del uso de cada uno de los componentes de los servidores (de computación y



de datos) pertenecientes a CVBLab. Finalmente, en la tabla 1.4 se calcula el coste total en concepto del *hardware* utilizado.

Descripción	Cantidad	Coste unitario (€) sin IVA	Periodo de amortización (meses)	Intervalo amortizado (meses)	Coste imputable (€) sin IVA
ASUS M51AC-IT012S intel Core i7-4770, 3.4 GHz, 8 GB RAM, 32 GB, nVIDIA GeForce GTX 760"	1	882,15	72	8	98,02
<b>TOTAL</b>					<b>98,02 €</b>

**Tabla 1.2:** Coste del uso del equipo personal.

Descripción	Cantidad	Coste unitario (€) sin IVA	Periodo de amortización (meses)	Intervalo amortizado (meses)	Coste imputable (€) sin IVA
procesador Intel i7 @4.20GHz	1 u	344,00	48	1	7,16
Tarjeta gráfica NVIDIA Titan V	1 u	3300,00	48	1	68,75
Disco SSD de 250 GB	1 u	77,00	48	1	1,60
<b>TOTAL</b>					<b>77,51 €</b>

**Tabla 1.3:** Coste del uso de los servidor del grupo CVBLAB.

Descripción	Coste imputable (€) sin IVA
ASUS M51AC-IT012S intel Core i7-4770, 3.4 GHz, 8 GB RAM, 32 GB, nVIDIA GeForce GTX 760"	98,02
Servidor computación CVBLab	77,51
<b>TOTAL</b>	
<b>175,53 €</b>	

**Tabla 1.4:** Descomposición de los costes de hardware.

### 1.2.3 Costes de software

A continuación se exponen los costes necesarios para la obtención de las licencias de software imprescindibles para el desarrollo del proyecto. El ordenador utilizado incluye la licencia del sistema operativo Windows 10, así como sus posteriores actualizaciones. Por otro lado, la redacción del documento se ha llevado a cabo mediante el editor de texto online *www.overleaf.com*, de uso gratuito. Las librerías cuDNN utilizadas también son gratuitas debido a que son de código abierto. Si que ha supuesto un coste la utilización de la versión de Matlab empleada para el desarrollo de este proyecto. En la tabla 1.5 se puede observar el coste del software utilizado.

Descripción	Cantidad	Coste (€) sin IVA	Periodo de amortización (meses)	Intervalo de uso (meses)	Coste imputable (€) sin IVA
Overleaf	1	0	12	8	0
Matlab R2018a	1	800	12	8	533,3
Librerías cuDNN	1	0	12	1	0
<b>TOTAL</b>					<b>533,3 €</b>

**Tabla 1.5:** Descomposición de los costes de software. Todos los precios corresponden a las licencias de **Uso Académico**.

## 1.3 Presupuesto Total

Una vez calculado el coste relativo al personal, hardware y software se computa el presupuesto de Ejecución de Material (PEM) que es la suma de los costes anteriormente citados. Esta operación puede verse ejemplificada en la Tabla 1.6.

Descripción	Coste (€)
Costes de mano de obra	5.750,00
Costes de hardware	175,53
Costes de software	533,3
<b>Suma (PEM)</b>	<b>6.458,3</b>

**Tabla 1.6:** Presupuesto de Ejecución de Material del proyecto.

Posteriormente, se calcula el importe total que supondría el desarrollo de este Trabajo Fin de Grado incorporando a los costes de Ejecución de Material la parte proporcional a los costes generales y al beneficio industrial. Además, se incluyen los impuestos y otros gastos. Concretamente, se añade un 13 % correspondiente a los gastos generales y un 6 % en concepto del beneficio industrial. Por último, a este resultado se le añade un 21 % adicional correspondiente al Impuesto sobre el Valor Añadido (IVA). El importe total se detalla en la Tabla 1.7.

Descripción	Coste (€)
Presupuesto de Ejecución de Material	6.458,3
Gastos generales	839,58
Beneficio industrial	387,5
<b>Suma</b>	<b>7.685,38</b>
IVA (21 %)	1.613,93
<b>PRESUPUESTO TOTAL</b>	<b>9.299,31 €</b>

**Tabla 1.7:** Presupuesto total del presente proyecto.

Por tanto, el coste total de la realización de este proyecto asciende a **NUEVE MIL DOSCIENTOS NOVENTA Y NUEVE EUROS CON TREINTA Y UN CÉNTIMOS**.