UNIVERSITAT POLITÈCNICA DE VALÈNCIA

_ TELECOM ESCUELA TÉCNICA VLC SUPERIOR DE INGENIERÍA DE TELECOMUNICACIÓN

IDENTIFICACIÓN AUTOMÁTICA DE GLAUCOMA A PARTIR DE IMÁGENES OCT CIRCUMPAPILARES MEDIANTE EL USO DE REDES NEURONALES CONVOLUCIONALES

Sheyla Nieves del Amo

Tutor: Valery Naranjo Ornedo

Cotutor: Adrián Colomer Granero

Trabajo Fin de Grado presentado en la Escuela Técnica Superior de Ingenieros de Telecomunicación de la Universitat Politècnica de València, para la obtención del Título de Graduado en Ingeniería de Tecnologías y Servicios de Telecomunicación

Curso 2018-19

Valencia, 3 de julio de 2019

Escuela Técnica Superior de Ingeniería de Telecomunicación
Universitat Politècnica de València
Edificio 4D. Camino de Vera, s/n, 46022 Valencia
Tel. +34 96 387 71 90, ext. 77190
www.etsit.upv.es

VLC/ CAMPUS VALENCIA, INTERNATIONAL CAMPUS OF EXCELLENCE

EMAS Gestión ambiental verificada REG.NO. ES-CV- 000030

A Valery y Adrián, por introducirme en el apasionante mundo de la inteligencia artificial e ingeniería biomédica.

A mi familia y amigos, por el apoyo incondicional mostrado durante toda mi etapa académica.

Y especialmente, a Lidia y José Luis por su ayuda y cariño ofrecidos durante estos meses.

RESUMEN

El glaucoma es una de las patologías que ocasionan mayor pérdida de visión e incluso ceguera en una parte considerable de la población. Una de las mayores limitaciones que se presenta en la práctica clínica, es la dificultad de detección y diagnóstico de dicha patología en las primeras etapas de su formación. Por tanto, resulta esencial disponer de buenas técnicas de detección temprana. En este contexto, la tomografía de coherencia óptica realiza grandes aportaciones, ya que proporciona imágenes de las capas de la retina y medidas referentes a los espesores de dichas capas. Las exploraciones centradas en el nervio óptico (circumpapilares) son las más utilizadas, ya que es en el nervio donde es más apreciable dicha patología.

Así pues, el objetivo perseguido en este trabajo fin de grado es la creación de modelos de predicción que, partiendo de imágenes circumpapilares, permitan discernir entre ojos sanos y glaucomatosos. Concretamente, se recurrirá al empleo de técnicas de inteligencia artificial basadas en aprendizaje profundo, tales como las redes neuronales convolucionales dado que, en el entorno clínico, han resultado clave para el diagnóstico de imágenes por computador.

Palabras clave: *inteligencia artificial, aprendizaje profundo, redes neuronales convolucionales, transferencia de conocimiento, clasificación, glaucoma, tomografía por coherencia óptica.*

RESUM

El glaucoma és una de les patologies que ocasionen major pèrdua de visió i inclús ceguera en una part considerable de la població. Una de les majors limitacions que es presenta en la pràctica clínica, és la dificultat i diagnòstic d'aquesta patologia a les primeres etapes de la seua formació. Per tant, resulta essencial disposar de bones tècniques de detecció primerenca. En aquest context, la tomografia de coherència òptica realitza grans aportacions, ja que proporciona imatges de les capes de la retina i mesures referents a les grossàries d'aquestes capes. Les exploracions centrades al nervi òptic (circumpapilars) són les més utilitzades, ja que és al nervi on és més apreciable dita patologia.

Així doncs, l'objectiu perseguit en aquest treball fi de grau és la creació de models de predicció que, partint d'imatges circumpapilars, permeten discernir entre ulls sans i ulls amb glaucoma. Concretament, es recorrerà a l'emprament de tècniques d'intel·ligència artificial basades en aprenentatge profund, com ara les xarxes neuronals convolucionals atés que, a l'entorn clínic, han resultat clau per al diagnòstic d'imatges per computador.

Paraules clau: *intel·ligència artificial, aprenentatge profund, xarxes neuronals convolucionals, transferència de coneixement, classificació, glaucoma, tomografia per coherència òptica.*

ABSTRACT

Glaucoma is one of the pathologies that causes the greatest loss of vision and even blindness among a significant part of the population. A major limitation of clinical practice is the difficulty of detecting and diagnosing this pathology in the early stages of its formation. Therefore, good early techniques are essential. In this context, optical coherence tomography offers great contributions, as it provides images from the retina layers and measurements referring to the thickness of these layers. The examinations focused on the optic nerve (circumpapillary) are the most used, since it is the nerve where this pathology is more visible.

So the aim of this final degree work is the creation of prediction models that, based on circumpapillary images, allow discerning between healthy or glaucomatous eyes. Specifically, artificial intelligence techniques based on deep learning will be used, such as convolutional neural networks, since they have been essential for computer imaging diagnosis in the clinical scenario.

Keywords: *artificial intelligence, deep learning, convolutional neural networks, transfer learning, classification, glaucoma, optical coherence tomography*.

ÍNDICE

Índice de figuras

Índice de gráficas

Índice de tablas

Capítulo 1. Introducción

En los últimos años, el área de las tecnologías de telecomunicación ha experimentado una gran evolución que conlleva a la inclusión de las TIC en diversas áreas y sectores económicos. Concretamente, la rama especializada en el análisis de imágenes y procesado de señales está adquiriendo especial relevancia en el ámbito clínico y de la medicina, dado su potencial para el diagnóstico por imagen.

La medicina es una de las ciencias que requiere de mayor precisión y fiabilidad. Para ello, resulta necesario evaluar paralelamente gran variedad de datos referentes a los historiales clínicos de los pacientes, así como tener presente gran cantidad de información médica, diagnósticos, métodos de detección, antecedentes familiares y casos similares a los del sujeto en cuestión para determinar el diagnóstico de cada paciente nuevo. Este hecho hace necesario la automatización de tareas.

Por otro lado, el creciente número de datos diarios junto con la necesidad de encontrar modelos predictivos de mayor precisión que los actuales y de manera automática, ha dado lugar a la expansión de la inteligencia artificial en todas las áreas de trabajo. El uso de la inteligencia artificial, entre otras utilidades, facilita el diagnóstico por imagen gracias a la irrupción de novedosas técnicas que, combinadas de forma idónea aprenden a extraer las características más relevantes de las imágenes para discernir entre pacientes sanos y patológicos.

De esta forma, la cooperación conjunta entre médicos, ingenieros de telecomunicación y biomédicos con alto dominio y habilidades en el entorno de la inteligencia artificial pueden contribuir a la consecución de mejores sistemas de diagnóstico.

En este contexto surge Galahad [1], un proyecto europeo implantando dentro del Horizonte 2020 [H2020-ICT-2016-2017, 732613] en el que diversas entidades e instituciones participan aportando conocimiento tanto en el área de la medicina como en lo puramente tecnológico, para la consecución de un objetivo común. Entre ellos, participa el *Computer Vision and Behaviour Lab (CVBLab)*, de la Universidad Politécnica de Valencia.

Galahad es una iniciativa en la que se investigan sistemas de detección temprana de glaucoma basadas en tomografía de coherencia óptica de bajo coste y de mayor resolución para su posible instauración en centros de atención primaria. El principal inconveniente que presenta el glaucoma es que la pérdida de visión, por lo general, suele experimentarse de forma lenta y progresiva por lo que es difícil de detectar hasta alcanzar etapas avanzadas. Por este motivo, disponer de herramientas que permitan identificar la patología en épocas tempranas es esencial.

Investigaciones recientes demuestran que uno de los métodos más eficientes para la detección de dicha patología es la medición del espesor de las capas de la retina, a la entrada del nervio óptico ya que, en su formación, el glaucoma genera un aumento en la excavación de células ganglionares que repercute disminuyendo el espesor de la fibra del nervio. Sin embargo, un inconveniente asociado a esto último es que en muchas ocasiones la segmentación de las capas ha de hacerse manualmente, lo que conlleva gran inversión de tiempo. Además, es necesario disponer de diversos datos clínicos como son la presión intraocular del paciente y su campo visual, entre otros para la correcta identificación.

En este contexto, la OCT proporciona una solución tecnológica aportando imágenes a partir de las cuales se puede obtener dicho espesor. Con este trabajo de fin de grado, se pretende hacer uso de las imágenes circumpapilares proporcionadas por este sistema para tratar de encontrar modelos de clasificación automáticos basados en inteligencia artificial que permitan detectar la existencia de glaucoma. Así, se pretende obtener altas tasas de precisión que permitan disminuir el número de falsos positivos y negativos al mínimo posible.

Capítulo 2. Objetivos

El principal objetivo de este trabajo fin de grado consiste en el desarrollo de modelos de clasificación automática basados en aprendizaje profundo que, a partir del empleo de imágenes OCT circumpapilares permitan diferenciar entre pacientes sanos y pacientes que padecen glaucoma. Para ello, se han fijado una serie de propósitos que serán la guía de este proyecto:

- Estudio de la anatomía ocular en general y del glaucoma en particular de forma que se adquiera el conocimiento suficiente en lo que respecta a la formación de dicha patología, que proporcionará los conceptos necesarios para conformar una opinión crítica en base al tratamiento de las imágenes circumpapilares disponibles, así como en la posterior interpretación de los modelos diseñados.
- Investigación de técnicas existentes en el ámbito clínico para la detección de glaucoma, así como de modelos que ya han sido propuestos por otros autores tomando como punto de partida técnicas basadas en inteligencia artificial.
- Diseño, desarrollo e implementación de una arquitectura neuronal convolucional propia mediante el empleo de diferentes bloques convolucionales. Tras ello, diversas modificaciones serán propuestas con el objetivo de obtener mejores resultados de clasificación.
- Hacer uso de arquitecturas pre-entrenadas complejas mediante la técnica de transferencia de conocimiento, de forma que permita realizar una comparativa con el modelo diseñado desde cero, valorando las diferentes ventajas e inconvenientes que ello supone.
- Elección del modelo de clasificación que permita discernir mejor entre pacientes sanos y glaucomatosos atendiendo a diversas métricas y criterios.

Capítulo 3. Metodología

Para la elaboración de este trabajo de fin de grado, diversas tareas han sido llevadas a cabo, las cuales pueden ser distribuidas en cuatro fases atendiendo a la distribución temporal de las mismas.

En una etapa inicial, es necesario adquirir conocimiento en lo referente a la inteligencia artificial, comprender los algoritmos que la definen y aprender a diferenciar entre las distintas técnicas disponibles. En esta fase inicial, también resulta conveniente introducirse en el lenguaje de programación *Python*, mediante el cual se implementarán los modelos propuestos.

En una segunda fase, se deberá llevar a cabo la búsqueda de información que engloba el marco teórico del trabajo, es decir, el glaucoma, la herramienta de detección OCT y diferentes técnicas de inteligencia artificial que puedan dar respuesta al problema de detección.

Cuando todas estas tareas concluyen comienza la fase de desarrollo de modelos, donde diferentes propuestas serán implementadas y entrenadas para, finalmente, proceder a compararlas analizando las ventajas e inconvenientes que cada una de ellas presenta.

Finalmente, nos encontramos en disposición de realizar la memoria final del documento.

A continuación se muestran enumeradas las diversas tareas llevadas a cabo en cada una de las fases anteriormente mencionadas.

1. Trabajo previo:
   - Proceso de aprendizaje previo para la obtención de conocimiento en el área de la inteligencia artificial y el *machine learning* mediante diversas herramientas, entre las que se encuentra la realización de cursos de introducción a la inteligencia artificial.
   - Aprendizaje del lenguaje de programación *Python* necesario para la programación de los modelos de clasificación que van a ser desarrollados.

2. Formación teórica y preparación de la base de datos:
   - Búsqueda de información acerca del glaucoma.
   - Estudio de diferentes herramientas empleadas para la detección de la patología.
   - Revisión bibliográfica sobre modelos de detección de glaucoma mediante el empleo de técnicas basadas en inteligencia artificial.
   - Elaboración de la base de datos.

3. Desarrollo de modelos:
   - Diseño e implementación de una arquitectura neuronal convolucional diseñada desde cero.
   - Aplicación de la técnica de aumento sintético de datos a la arquitectura anteriormente diseñada.
   - Introducción de variables demográficas tales como la edad y el sexo al modelo diseñado.
   - Elaboración de modelos basados en arquitecturas ya existentes, haciendo uso de la técnica de transferencia de conocimiento.
   - Comparativa entre diferentes arquitecturas y extracción de conclusiones.

4. Redacción y revisión de la memoria final del proyecto.

La distribución temporal en la que dichas tareas han sido desarrolladas se muestra en el Anexo I. Para finalizar, cabe añadir que dado que se requiere partir de una base sólida de conocimiento basado en inteligencia artificial e implementación de modelos mediante *Python*, estas tareas han sido confeccionadas a lo largo de todo el curso académico 2018/2019.

Capítulo 4. Marco teórico

El glaucoma es una patología responsable de gran parte de casos de ceguera en el mundo. Además, la mayoría de los casos se producen de forma lenta, progresiva y asintomática, por lo que resulta difícil de detectar en etapas tempranas. Su principal inconveniente es que carece de cura, y si bien es posible controlar su evolución. Por este motivo, resulta esencial disponer de sistemas de detección temprana del mismo. En este capítulo, se pretende contextualizar al lector en el marco teórico de la anotomía ocular y más concretamente en el glaucoma así como en la tecnología de detección OCT, instrumento que proporciona las imágenes que permitirán el desarrollo este trabajo de fin de grado.

4.1 Anatomía del ojo

4.1.1 Concepto de visión

El ojo es el órgano que facilita sentido de la vista. Su especialidad consiste en transformar estímulos luminosos en estímulos nerviosos, los cuales son transmitidos a través del nervio óptico hasta el cerebro, donde finalmente se construye la imagen.

La visión es un proceso complejo que consta de varias etapas. Comienza con la organización del estímulo nervioso procedente de la refracción de los rayos luminosos y el enfoque de las imágenes sobre la retina. Gracias a los fotorreceptores situados en la retina, se produce una fototransducción, convirtiendo los fotones captados en una señal nerviosa. A continuación, esta señal es transmitida a través del nervio óptico hasta el tálamo[1]. En él, se produce una amplificación de la señal y se elimina la información no relevante. Esta señal llega hasta el córtex, donde se produce la decodificación de la señal visual. Finalmente se produce una retroalimentación en el sistema visual, generándose reflejos tales como la acomodación, graduación de la abertura pupilar o control de movimientos oculares [2].

4.1.2 Estructura del globo ocular

En la Figura 1 se muestra la representación anatómica del globo ocular, enumerada con las estructuras relevantes que serán descritas a continuación, de modo que sirva para facilitar la comprensión al lector.

[1] El tálamo es una parte del encéfalo cuya misión principal consisten en la regulación de los sentidos.

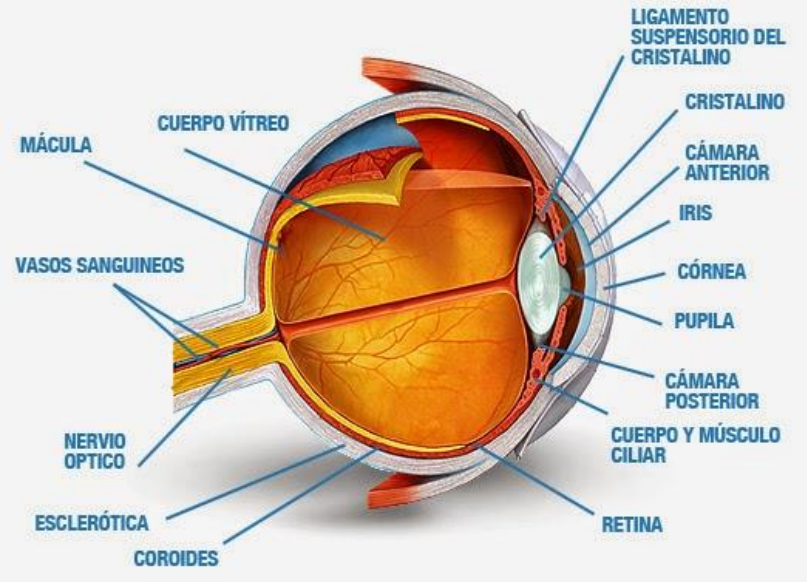


Figura 1: Anatomía del ojo. Enumeración de las diferentes estructuras que conforman el globo ocular [3].

El globo ocular se encuentra dividido en tres capas: la túnica fibrosa, la túnica vascular y la túnica interna nerviosa.

1. *Túnica fibrosa:*

Es la encargada de dar consistencia al ojo, por lo que está representada por la esclerótica, la córnea y los humores acuoso y vítreo. La esclerótica es una membrana resistente, de naturaleza fibrosa, que envuelve cuatro quintos de la parte posterior del globo ocular. Su función es proteger a los elementos internos del globo ocular y mantener el tono ocular para conservar el volumen y la forma del ojo.

La córnea es una membrana fibrosa, transparente y refringente que se encuentra en el polo anterior del globo ocular, formando un ángulo agudo con la esclerótica, lo cual favorece la convergencia de los rayos luminosos sobre la pupila. Su transparencia es debida a la ausencia de vasos sanguíneos, a la orientación paralela de sus fibras y al grado de hidratación de la misma. Presenta inervación sensible por lo que permite el parpadeo. Además tiene función de protección y consistencia del globo ocular, y de canalizar la luz hacia el cristalino.

El cristalino es una lente transparente, biconvexa y flexible que se encuentra suspendida entre el iris, la pupila y el humor vítreo. Su cara posterior presenta una curvatura mayor que la zona anterior, lo que hace que la luz converja sobre la retina. Es el responsable de enfocar los rayos luminosos de modo que sea capaz de formarse una imagen correcta sobre la retina, es decir, es el responsable del proceso de acomodación del ojo. Así, si se encuentra relajado permite enfocar objetos a gran distancia mientras que si aumenta su curvatura, se enfocan objetos cercanos [4].

El ojo, además, cuenta con una serie de cámaras o espacios de gran importancia: el humor acuoso (cámara anterior y posterior) y el cuerpo vítreo (cámara vítrea).

El cuerpo o humor vítreo es un fluido gelatinoso que se encuentra rodeado por una membrana limitante llamada hialoides. Ocupa la cámara vítrea, en concreto, situado entre el cristalino y la retina. Su función principal es proporcionar consistencia al globo ocular manteniendo su forma.

El humor acuoso es un líquido transparente, que ocupa las cámaras anterior y posterior del globo ocular, entre la córnea y el cristalino. Se genera por los procesos ciliares, desde donde se vierte a la cámara anterior del ojo a través de la pupila. Su función consiste en seguir un ciclo de regeneración, recorrido y drenaje de suma importancia, ya que una obstrucción del drenaje conduce a hipertensión ocular y glaucoma. Una vez en la cámara anterior, continúa su recorrido de drenaje adentrándose a la cámara posterior a través de la malla trabecular situada en el ángulo iridocorneal (el cual se sitúa entre el iris y la córnea). A continuación atraviesa el conducto de Schlemm o seno venosos de la esclerótica, donde finalmente es eliminado.

Tanto la córnea como el cristalino son estructuras no vascularizadas, por lo que su modo de nutrición y fuente de oxígeno es el humor acuoso.

2. Túnica vascular

La túnica vascular está conformada por tres componentes principales. El primero de ellos es la coroides, una capa vascular que se encuentra entre la retina y la esclerótica.

El iris es un disco circular con una abertura central conocida como pupila. La pupila actúa como un diafragma, regulando la cantidad de luz que entra en el globo ocular. Además separa la cámara anterior de la cámara posterior del ojo.

Finalmente, el cuerpo ciliar tiene dos componentes, uno de ellos es el músculo ciliar, que se une al cristalino mediante fibras, y unos pliegues "procesos ciliares" productores del humor acuoso.

3. Túnica interna/nerviosa

La retina es la túnica nerviosa del globo ocular, responsable de la parte sensible del ojo. Se trata de una membrana fina y transparente que se relaciona con la coroides en la cara exterior, y con el humor vítreo en la cara interior.

Los fotorreceptores son neuronas especializadas que contienen pigmentos fotoquímicos capaces de reaccionar ante la luz. Existen dos tipos de fotorreceptores:

- Los conos que son capaces de reaccionar a distintas longitudes de onda, hecho que permite realizar la distinción de colores.
- Los bastones que reaccionan únicamente a escala de grises, de modo que su función es definir formas.

Estos fotorreceptores se encuentran en contacto, por un lado, con la capa pigmentaria de la retina no visual, la cual se encuentra junto con la coroides y, por otro lado, con una membrana limitante externa. A esta le siguen una serie de capas que en último término contactan con la capa de células ganglionares y los axones de estas células, que constituyen el nervio óptico. La localización exacta de los receptores dentro de la retina se puede encontrar en la Figura 2.

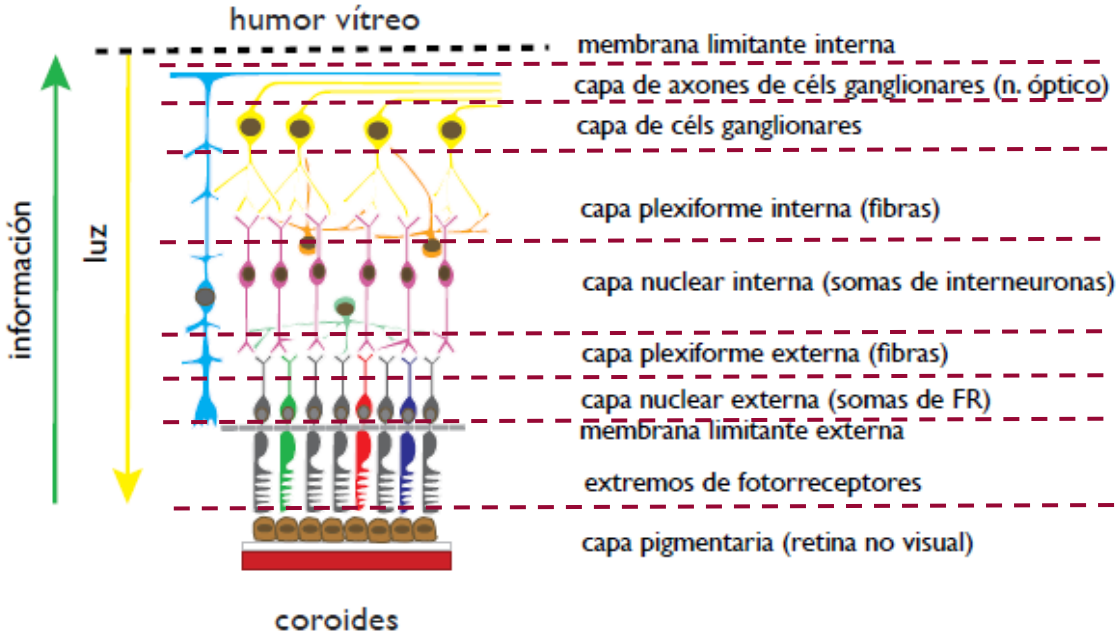


Figura 2: Localización de los fotorreceptores y diferentes capas de la túnica nerviosa, previas a la formación del nervio óptico.

La papila o punto ciego[2] de la retina se encuentra en la región posterior de la misma, donde los axones de las células ganglionares confluyen para formar el nervio óptico. Es en esta zona donde entran los vasos sanguíneos.

La fóvea o mácula amarilla es una región muy sensible que permite la visión aguda gracias a una mayor densidad de fotorreceptores. En el centro de la mácula se encuentra la foveola, formada únicamente por conos especializados que son más sensibles que el resto. Dicha sensibilidad también se debe a que en esta zona cada cono conecta exclusivamente con una célula bipolar y una horizontal.
Una vez conocida la estructura del globo ocular, nos adentraremos en la definición del glaucoma, donde el humor acuoso supone un papel fundamental.

[2] Su nombre, "punto ciego", es debido a que en él no hay fotorreceptores.

4.2 Glaucoma

4.2.1 Definición, factores de riesgo y epidemiología

Definición

El glaucoma consiste en un conjunto de enfermedades oculares que suelen estar caracterizadas por el aumento de la presión intraocular dentro de la cámara anterior del ojo, ocasionándose una pérdida de células ganglionares de la retina y daños en el nervio óptico. Así pues, se trata de una neuropatía óptica degenerativa crónica e irreversible, que lleva a la ceguera [5].

La presión intraocular normal en adultos, o PIO, adquiere valores de media entre 15±2.5 mmHg, cifras que pueden variar a lo largo del día. Cuando esta PIO supera los 21 mmHg durante un periodo de tiempo amplio y producen daños en las células ganglionares se origina el glaucoma. Es muy importante destacar que es imprescindible que exista daño del nervio óptico para poder hablar de glaucoma, ya que puede haber casos de glaucoma con una tensión ocular reducida; y presión intraocular por encima de los 21 mmHg sin que se presenten indicios de glaucoma [4].

Esta tensión ocular está determinada por el equilibrio entre la tasa de secreción y drenaje del humor acuoso. El humor acuoso es un líquido generado en los procesos ciliares y después segregado a la cámara posterior, cuya función principal consiste en nutrir a la córnea y al cristalino. Este humor fluye desde la cámara posterior a la cámara anterior a través de la pupila. Para ello, ha de abordar el camino entre el iris y el cristalino. Sin embargo, el iris es plano y se encuentra situado en la cámara anterior del cristalino, por lo que el humor acuoso necesita presión suficiente para vencer la resistencia impuesta por el iris, despegarlo del cristalino y poder fluir entre ambos. Si la resistencia al flujo del humor acuoso aumenta, se incrementará también la presión intraocular.

Una vez el humor acuoso ha alcanzado la cámara anterior, utiliza dos vías diferentes para salir de ella:

- Vía trabecular: es la principal ya que el 90% del humor acuoso se drena a través de la malla trabecular hasta alcanzar el canal de Shlemm. Esta vía es muy sensible a la presión intraocular.
- Vía uveoescleral: drena el 10% de humor acuoso restante. El humor acuoso de la cámara anterior fluye desde el sistema vascular uveoescleral para alcanzar la zona venosa general.
- Parte del humor acuoso también puede ser drenado directamente desde el iris.

En la Figura 3 puede observarse el recorrido llevado a cabo por el humor acuoso, así como las diferentes vías y resistencias que encuentra en el camino.

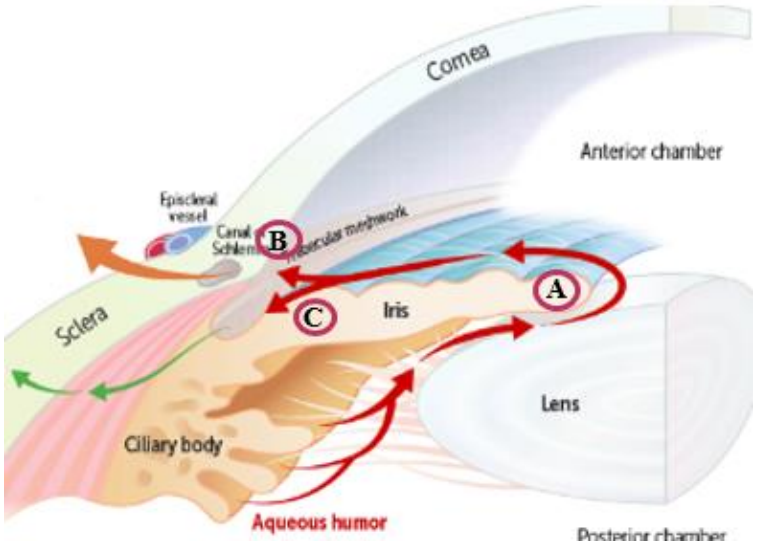


Figura 3: Fisiología de la circulación del humor acuoso. A: resistencia impuesta por la pupila B: vía trabecular, C: vía uveoescleral

De este modo, el glaucoma no se produce por un aumento de la generación de humor acuoso sino por la resistencia impuesta al flujo de salida de dicho humor, que conlleva un aumento de la presión intraocular. Los fallos de drenaje suelen ser ocasionados por la presencia de alguna obstrucción en las vías de drenaje del humor acuoso, que generalmente guardan relación con factores como la edad del paciente, la presencia de hematíes coagulados producidos en accidentes o por determinadas infecciones oculares.

Factores de riesgo

En cuanto a los factores de riesgo de sufrir glaucoma, resalta principalmente el aumento de la presión intraocular del paciente; pero otros factores se han de tener en cuenta como son la edad, el sexo e incluso la raza para determinados tipos de glaucoma, si el paciente en cuestión presenta hipermetropía, miopía o diabetes, así como antecedentes familiares.

Epidemiología

- El glaucoma es la segunda causa de ceguera más frecuente en los países desarrollados, después de la diabetes mellitus.
- Según la Organización Mundial de la Salud, el glaucoma es la segunda causa de ceguera en el mundo, después de las cataratas [6].
- La ceguera ocasionada a partir del glaucoma es entre 6 y 8 veces más común en afroamericanos que en caucásicos [7].
- De entre los casos de ceguera existentes, entre el 6.7-21% son producidos a causa del glaucoma.
- Lo padecen 45 de cada cien millares de habitantes mayores de 50 años, ampliándose esta cifra a 241 para mayores de 70.
- Se estima que existen aproximadamente un 50% de casos sin diagnosticar.

4.2.2 Tipos de glaucoma

Se puede establecer una clasificación del glaucoma atendiendo a tres categorías: glaucoma de ángulo abierto o crónico, glaucoma de ángulo cerrado o secundario y glaucoma congénito o infantil. Para las dos primeras categorías se puede definir una sub-clasificación en glaucoma primario y secundario, siendo la diferencia entre ambos que en los secundarios, el glaucoma no se genera por alteraciones en la anatomía del ojo como tal, sino que es causado por otras enfermedades oculares o factores como inflamaciones, traumatismos, hemorragias, tumores, o determinados medicamentos, entre otros.

1. Glaucoma de ángulo abierto o crónico:

   - Primario: Es el tipo de glaucoma más frecuente, suponiendo más del 90% de los casos. Es comúnmente conocido como "ceguera silenciosa" o "falsa miopía" ya que no genera síntomas hasta alcanzar fases avanzadas. Suele darse en pacientes que superan los 40 años de edad, con una PIO entre 20-35 mmHg. El motivo principal de su origen es el envejecimiento de la malla trabecular; cuando esto ocurre se vuelve más rígida por lo que pierde la capacidad de filtración, dando lugar al aumento de la presión intraocular. Todo ello conduce a la excavación del nervio óptico, produciéndose una ceguera progresiva irreversible comenzando desde la periferia, motivo por el cual el paciente puede tardar tiempo en percibir la pérdida de visión. Es por esta razón que se requieren sistemas de diagnóstico precoz para la detección y tratamiento del glaucoma en las primeras fases, antes de que el paciente perciba la pérdida irreversible de gran parte de su campo visual. En la Figura 4, se puede observar la disminución del campo visual en las diferentes fases de desarrollo del glaucoma.

Figura 4: Pérdida del campo visual en diferentes fases del glaucoma [8]

   - Secundario: En este caso, la malla trabecular no se encuentra afectada, pero otros agentes la obstruyen, lo que conlleva a un aumento de la resistencia a la salida del humor acuoso.

2. Glaucoma de ángulo cerrado
   - Primario: Este tipo de glaucoma es menos frecuente, supone el 6% de los casos de glaucoma. En este caso, la pérdida de visión se produce de forma repentina y, por lo general, dolorosa. El iris entra en contacto con la córnea, e incluso puede solaparse con la superficie anterior del cristalino, bloqueando así la cámara anterior, impidiendo de esta forma al humor acuoso salir de la cámara posterior. Este hecho provoca el aumento de la PIO hasta valores muy elevados, ente los 40-70 mmHg.
   - Secundario: similar al primario, pero producido por otros factores o enfermedades.

3. Glaucoma congénito o infantil: se da en uno de cada 15.000 nacimientos, siendo más probable en varones. Es producido por el aumento de la PIO de forma anómala en las primeras etapas de la vida de un recién nacido, que genera la dilatación de las paredes del globo ocular pero, sobre todo, de la córnea.

4.2.3 Métodos de detección

La detección y diagnóstico de glaucoma es un proceso complejo y costoso ya que, como se ha mencionado, no hay una causa única y concreta que dé origen a esta patología. Las técnicas empleadas actualmente se centran en el estudio del nervio óptico y de la presión intraocular, así como del campo visual del paciente. Sin embargo, ninguna de ellas por sí misma es capaz de determinar la presencia o no de glaucoma, aunque resulta relevante la aportación de cada una de ellas. Por este motivo, el oftalmólogo deber realizar diversas pruebas médicas y contrastar la información obtenida para poder dictaminar finalmente un diagnóstico.

Además, el especialista requiere de información adicional como es el historial clínico del paciente, datos demográficos y un test de agudeza visual, que analiza la capacidad de visión del ojo a distintas distancias.

A continuación, en la Tabla 1 se expone una breve descripción de los principales [4] sistemas actuales existentes para la detección de dicha patología.

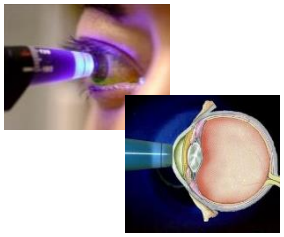
MÉTODO DE EVALUACIÓN		OBJETIVO	INSTRUMENTACIÓN	DESCRIPCIÓN
GONIOSCOPIA		Ángulo de drenaje		Se sitúa la cámara de tres lentes sobre el ojo. Si se aprecian todas las estructuras[3], el ángulo de la cámara está abierto, si no, cerrado.
TONOMETRÍA	Aplanación o de goldman	PIO normal		El tonómetro es colocado sobre la córnea y tratará de aplanarla. La PIO se determinará en función de la resistencia a la deformación impuesta por la córnea.
	Aire	PIO normal		Igual que la de aplanación, solo que en vez de aplanar la córnea, se le envía a esta una corriente de aire. Es menos precisa.
PAQUIMETRÍA		PIO real[4]		La resistencia que ofrece la córnea a la deformación guarda relación con su grosor. Consiste en medir dicho grosor para ajustar la PIO normal medida del paciente y obtener así la PIO real.
CAMPIMETRÍA		Campo Visual		Consiste en lanzar unos estímulos luminosos al paciente en diferentes posiciones espaciales, cada una de las cuales se corresponde con una fibra del nervio óptico. El paciente indica cuál de ellos percibe. Luego se mapean, y se obtiene un umbral mínimo de percepción.
TOMOGRAFÍA DE COHERENCIA ÓPTICA		Nervio óptico		Mide el hundimiento del nervio óptico, producido por el incremento de la presión intraocular. Supone, junto con la campimetría, la prueba más importante para la detección de glaucoma, ya que analiza el nervio óptico directamente.

Tabla 1: Métodos de detección y evaluación de glaucoma

[3] Las estructuras que se han de poder apreciar son el espolón escleral, la banda ciliar, la línea de Schwalbe pero, sobre todo y de especial relevancia el iris, la raíz del iris, la malla trabecular y la córnea.

[4] La presión intraocular normal hace referencia a la presión que ejerce el humor acuoso sobre las paredes internas del ojo. Sin embargo, cada sujeto posee un espesor de la retina diferente, lo que afecta directamente a la resistencia impuesta a la deformación de la misma. Así, la presión intraocular real consiste en ajustar la PIO normal al grosor de la retina de cada paciente.

4.3 Tomografía de coherencia óptica (OCT)

4.3.1 Definición

La Tomografía de Coherencia Óptica, habitualmente conocida como OCT, por las siglas de su terminología en inglés (*Optical Coherence Tomography*), es una técnica especialmente empleada en oftalmología para la detección y diagnóstico de enfermedades ocasionadas en la mácula y en la retina, además de suponer un instrumento muy útil para la detección y seguimiento del glaucoma. Se trata de un método no invasivo e indoloro que proporciona imágenes de alta resolución de la cabeza del nervio óptico, del vítreo y de la retina [9].

En el tema que nos ocupa, la OCT permite detectar y monitorizar el glaucoma, en combinación con otras técnicas e información del historial del paciente. Por una parte, proporciona imágenes transversales de la papila de forma radial, por lo que provee un análisis morfológico del nervio óptico[10]. Por otra parte, y a su vez, la funcionalidad de la OCT más empleada, es que permite visualizar la representación de las fibras nerviosas retinianas en forma de imágenes peripapilares. Concretamente, las imágenes muestran el área circular situada a 3.4 mm del nervio óptico. Dichas imágenes son muy utilizadas en la práctica para medir el espesor de las distintas capas del nervio, cuya disminución puede ser debida a diferentes enfermedades oculares, entre las que se encuentra el glaucoma.

En cuanto a sus orígenes, la idea de la OCT se basa en la anterior ecografía B, que utiliza ultrasonidos para la detección de enfermedades. La novedad que incorpora la OCT, es emplear haces de luz en lugar de ultrasonidos alcanzando mayor velocidad de trabajo ya que la velocidad de la luz es del orden de 10^6 la velocidad del sonido; además de suponer una resolución diez veces mayor a su antecesora ecografía B. De esta forma, se consigue la medición de estructuras y distancias en una escala de 10µm, en lugar que las 100µm del ultrasonido.

Su gran utilidad es que los dispositivos OCT disponen de un *software* embebido que permite comprar el espesor de las capas del paciente en cuestión con los datos normales que presentaría un conjunto de pacientes sanos de la misma edad, obtenidos a partir de una gran base de datos. Esta técnica, proporciona una sensibilidad y especificad del 90% en el diagnóstico y monitorización de la progresión del glaucoma [5].

Dada la alta resolución de las imágenes proporcionadas, estas permiten identificar estructuras retinianas de espesor reducido, de resolución 10-20µm tales como la membrana limitante externa y la capa de células ganglionares. Además, determinados *softwares* permiten visualizar cuantitativamente el grosor de las capas de la retina y un mapa de colores topográfico de colores falsos. Nosotros trabajaremos con imágenes OCT en escala de grises.

4.3.2 Principio de funcionamiento

La OCT, basa su modo de funcionamiento en el interferómetro de Michelson, que utiliza luz de coherencia baja en el espectro de infrarrojo, con una longitud de onda de 820-830 nm, en banda ancha. Un interferómetro es un instrumento que emite haces de luz de la misma longitud de onda, y mide las interferencias generadas en las ondas emitidas. Está compuesto por un divisor de haces, un frente de luz (normalmente un láser de diodo), un espejo de referencia y un detector. Concretamente, consiste en la emisión de dos rayos de luz propagándose en direcciones perpendiculares, de modo que cada haz recorrerá una trayectoria diferente, se reflejará y, finalmente, ambos convergen en un punto donde se mide la interferencia. En este momento, puede producirse interferencia constructiva o destructiva, dependiendo si los rayos recibidos se encuentran en fase o en contrafase, respectivamente. La onda resultante de la recombinación de ambos rayos de luz se envía al detector, donde finalmente se mide la potencia [11].

En el caso de aplicar el interferómetro a la OCT, el haz de luz emitida se divide en dos vías, una de ellas se dirige directamente al tejido ocular que se quiere analizar, mientras que el otro haz será enviado a un espejo de referencia situado a una distancia conocida. Esto permite conocer la posición de las diferentes estructuras oculares ya que cada una proporcionará un retardo de reflexión diferente con respecto al espejo de referencia.

De este modo, la imagen final que proporciona la OCT no será más que la combinación de la intensidad de la luz de referencia reflejada junto con la intensidad de la reflectividad de los tejidos oculares. La interferencia más intensa será producida por los tejidos que reflejan más luz. Las estructuras de alta reflectividad son representadas por el color rojo, las de reflectividad intermedia en verde y amarillo, mientras que las de baja reflectividad, en azul y negro. Del mismo modo, se puede obtener la imagen en escala de grises, siendo en este caso el color negro los tejidos de menor reflectividad.

Finalmente, se realiza una gráfica con las diferentes distancias recibidas y se obtiene una imagen en sentido axial, conocida como A-scan. Así, tras implementar un conjunto de imágenes axiales, se puede construir una imagen bidimensional (B-scan), que será la que el oftalmólogo visualizará para establecer su diagnóstico. En la Figura 5 se puede apreciar el modo de funcionamiento del interferómetro.

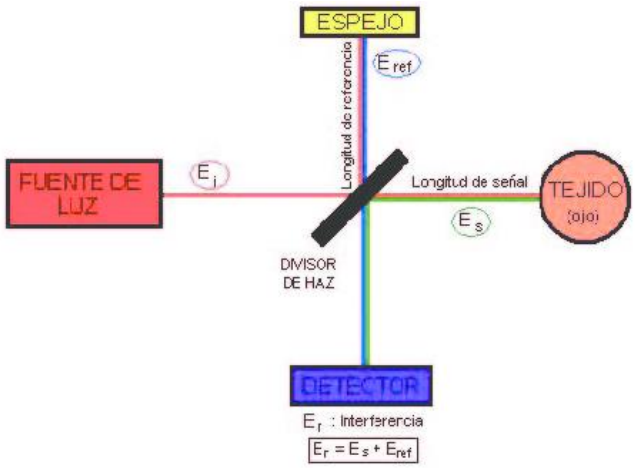


Figura 5: Representación gráfica del interferómetro de Michelson.

Capítulo 5. Modelos predictivos para la clasificación automática de glaucoma.

En este capítulo, se va a proceder a definir el concepto de inteligencia artificial, así como su utilidad en la tarea de clasificación automática. Seguidamente, se detallarán el tipo de arquitecturas de aprendizaje profundo de las que se hará uso en el próximo capítulo para la creación de modelos predictivos de detección de glaucoma, junto con una explicación del proceso de entrenamiento en un paradigma basado en redes neuronales. Finalmente, se hará una breve revisión bibliográfica de los avances realizados hasta el momento referidos a la identificación automática de glaucoma mediante el uso de aprendizaje profundo e imágenes circumpapilares.

5.1 Introducción

En los últimos años, con el avance de la investigación y la innovación, se está llevando a cabo la inclusión de la inteligencia artificial en diversas áreas y sectores de la economía. El nacimiento de la inteligencia artificial parte de orígenes remotos, de la mano de psicólogos, matemáticos e ingenieros cuya ambición consistía en intentar explicar el razonamiento y funcionamiento de la mente, para poder reproducirlo posteriormente en máquinas. Tras los esfuerzos realizados por un gran número de investigadores, el término de inteligencia artificial surge formalmente en 1956 en el congreso de Darmounth, propuesto por John McCarthy, profesor de la universidad de Stanford [12]. Desde entonces, la inteligencia artificial ha evolucionado a gran velocidad, dando origen a numerosas técnicas que facilitan el desarrollo de sistemas inteligentes que permiten la automatización de tareas. Algunos ejemplos de ello son los sistemas de detección de objetos [13] y clasificación automáticos, técnicas de reconocimiento facial y análisis de emociones [14], el procesado natural de lenguaje [15] y traducción automática, y un largo etcétera.

Los sistemas de clasificación inteligentes son capaces de proporcionar mayor precisión que los métodos tradicionales basados en la estadística y solventar determinados problemas como pueden ser las no linealidades. Además, dado que permiten analizar grandes bases de datos en un tiempo reducido, proporcionan altas precisiones y facilitan la generalización a diferentes bases de datos.

Trasladando la técnica de clasificación al ámbito clínico y de la salud, existen grandes limitaciones para la clasificación manual de patologías. Por una parte, se requiere de la recopilación de grandes bases de datos y del etiquetado manual por parte de los médicos, lo cual requiere emplear mucho tiempo. Además, para la detección de una enfermedad o patología es preciso considerar diversos datos de un mismo paciente, todo su historial clínico así como de sus antecedentes, análisis de pacientes que presenten los mismos síntomas, y otra serie de datos demográficos[16]. Por este motivo, resulta conveniente la automatización de tareas.

En este contexto, el aprendizaje profundo ofrece grandes aportaciones. Concretamente, las redes neuronales son capaces de encontrar modelos y patrones entre amplias bases de datos. Dependiendo de la tarea de aprendizaje a resolver así como del tipo de variables involucradas, será conveniente emplear una red neuronal concreta, en función de cual sea el propósito de su aplicación. Así, cuando se trata de variables numéricas, el perceptrón multicapa puede suponer una gran alternativa de clasificación [17]. (Véase 5.2).

Del mismo modo, la inteligencia artificial puede favorecer gratamente el diagnóstico por imagen [18], [19]. De esta forma, dotando a los sistemas de numerosas imágenes etiquetadas que representen la estructura del cuerpo humano a analizar, tanto de pacientes sanos como de pacientes enfermos, se pueden obtener grandes clasificadores que ayuden a determinar si nuevos pacientes presentan o no la patología en cuestión. Concretamente, en un escenario representado por imágenes, una de las técnicas más eficiente son las redes neuronales convolucionales (Convolutional Neural Networks (CNNs)) [20], [21].

En caso de que el problema a resolver requiera de datos etiquetados, deberemos hacer frente a una situación de aprendizaje supervisado. En caso contrario, si los algoritmos de clasificación disponen de la habilidad de aprender y clasificar adecuadamente sin necesidad de conocer las etiquetas reales, (también denominadas *target* o *ground truth*), se tratará de un problema de aprendizaje no supervisado. Además, existe un híbrido entre ambos tipos de aprendizaje, el semi-supervisado, donde se conoce solamente el *ground truth* de algunos pacientes [16].

A continuación, se va a proceder a describir el fundamento teórico tanto del perceptrón como de las CNN, así como a la descripción de sistemas de clasificación que integran estas redes en su arquitectura.

5.2 El perceptrón multicapa

La red neuronal artificial recogida bajo el nombre de perceptrón multicapa (MLP del término *multi-layer perceptron* en inglés) procede del perceptrón simple, dada la necesidad de encontrar modelos potentes, capaces de hacer frente a conjuntos de datos que no presentan relaciones lineales[17].

Las redes neuronales pretenden ser un símil a la biología del cerebro humano para las máquinas inteligentes, de forma que la transmisión de conocimiento se produce a partir de neuronas interconectadas entre sí. De esta forma, cuando nueva información alcanza a una neurona, esta ha de procesarla, y, a continuación, transmitirla a las neuronas colindantes, de forma que se produzca la propagación de conocimiento. En la Figura 6 se muestra una imagen comparativa entre la anatomía neuronal biológica y la de una neurona artificial.

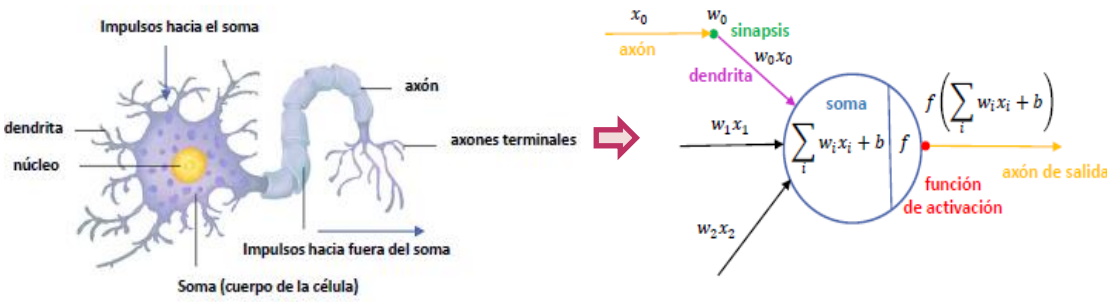


Figura 6: Comparativa ilustrativa entre la estructura de una neurona biológica y una neurona artificial.

Estas redes, están conformadas por tres tipos de capas. En primer lugar, se dispone de una capa de entrada, que estará compuesta por tantas neuronas como variables se deseen analizar. La capa de salida tendrá tantas neuronas como número de categorías en las que se quiera clasificar excepto que se trate de un modelo de clasificación binaria, en cuyo caso admite también el empleo de una única neurona binaria. Entre ambas, es posible encontrar una serie de capas, comúnmente conocidas como capas ocultas. Se dispondrán de tantas capas ocultas como profundidad se quiere que presente la red. Dichas capas serán las encargadas de ir transformando la información a la entrada en niveles superiores de abstracción con el objetivo de poder dividir los datos de la manera más precisa posible [22].

En función de la profundidad aportada, se tratará de una red neuronal poco profunda, conocida como *shallow neural network*, o de redes muy profundas, *deep neural network* atendiendo a su término en inglés.

Para transmitir la información entre las diferentes capas, se aplica una función de activación a la suma de los productos de las activaciones procedentes de las capas anteriores por sus pesos correspondientes, utilizando para ello el algoritmo detallado en la imagen de la derecha de la Figura 6.

La función de activación, también conocida como función umbral, puede adoptar diferentes formas. Algunos ejemplos de ello se muestran en la Tabla 2.

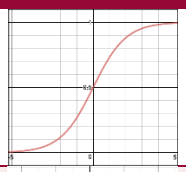
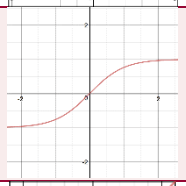
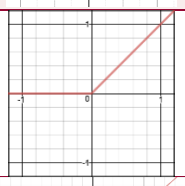
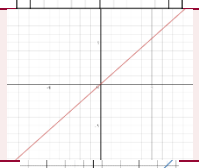
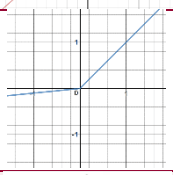
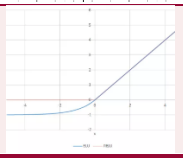
FUNCIÓN DE ACTIVACIÓN	EXPRESIÓN	CURVA	RANGO DE UMBRALIZACIÓN
Sigmoide	$f(x) = \frac{1}{1+e^{-x}}$		[0,1]
Tangente hiperbólica	$f(x) = \frac{2}{1+e^{-2x}} - 1$		[-1,1]
ReLU	$f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}$		[0,∞]
Lineal	$f(x) = cx$		[-∞,∞]
Leacky ReLU	$f(x) = \begin{cases} x, & x > 0 \\ \alpha x, & x \leq 0 \end{cases}$		[-∞,∞]
Elu	$f(x) = \begin{cases} x, & x > 0 \\ \alpha(e^x - 1), & x \leq 0 \end{cases}$		[-1,∞]

Tabla 2: Funciones de activación. Elaboración propia .Curvas tomadas de [21].

De esta forma, el MLP se adentrará en una fase de enteramiento conformada por dos etapas mediante las cuales se pretende dotar de aprendizaje a la red [23].

La primera de ellas se denomina propagación hacia delante, y consiste en transmitir la información de las entradas de la primera capa el resto de neuronas del modelo que, tras realizar las activaciones correspondientes llegarán a la capa de salida. En esta, y dado que se trata de aprendizaje supervisado, la salida del modelo es comparada con la etiqueta real de cada entrada y se calcula el error de predicción obtenido. En este momento da comienzo la fase de propagación hacia atrás, donde se actualizan los pesos de las diferentes neuronas tratando de encontrar el valor óptimo de los mismos.

Este proceso será llevado a cabo durante diversas épocas para que la red adquiera el conocimiento deseado.

Este proceso de aprendizaje puede ser observado en la Figura 7. En ella se muestran las diferentes capas de un MLP, junto con las entradas, pesos y salidas del modelo, así como dos fases que conforman el proceso de entrenamiento.

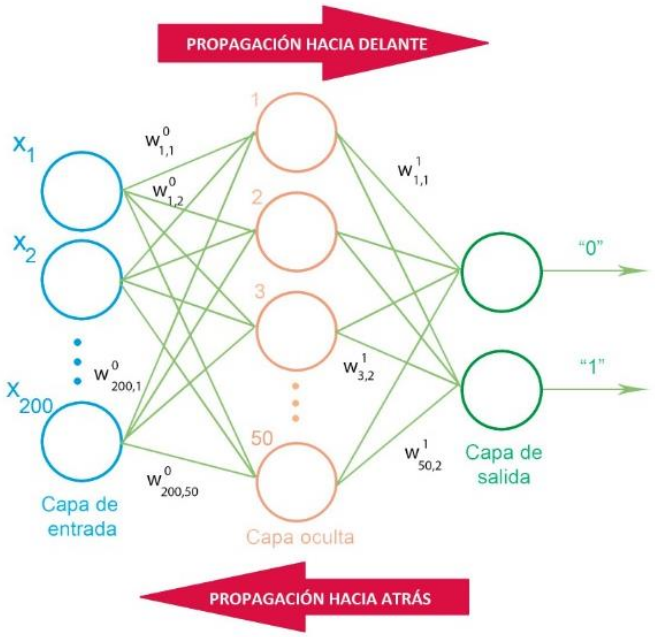


Figura 7: Arquitectura de un perceptrón multicapa compuesto de una única capa oculta destinado a la clasificación de dos categorías.

5.3 Redes Neuronales Convolucionales

Las redes neuronales convolucionales tienen su origen en 1995 de la mano de Yann LeCun et al., quienes propusieron una CNN para el reconocimiento automático de caracteres [24]. Las CNNs están basadas en el funcionamiento de la corteza visual del humano, permitiendo el reconocimiento de bordes y formas de los objetos, haciendo uso principalmente de dos estructuras: capas convolucionales y capas de agrupación.

Así pues, la entrada a la CNN son imágenes, normalmente en el espacio de color RGB. A continuación, cada imagen se introduce por un conjunto de filtros convolucionales y otras capas con el fin de extraer sus características esenciales para, seguidamente, realizar la clasificación de estas.

5.3.1 Arquitectura de las CNN. Definición de las capas.

- ***Capa convolucional***: Esta capa sirve para extraer características referidas a bordes de los objetos, relieves y curvas, entre otras. De esta forma, cuanto mayor nivel de exigencia se requiera de la red, mayor número de capas convolucionales se deberá incorporar a la arquitectura [25]. La capa convolucional se caracteriza por el empleo de filtros cuadrados, comúnmente conocidos como "*kernels*". Estos filtros recorren la imagen a clasificar con un determinado paso o *stride* y realizan una convolución (5.1) a los pixeles de la región que ocupa el filtro, que recibe el nombre de campo receptivo, dando lugar a un mapa de activación a su salida.

$$\boldsymbol{y[m,n] = x[m,n] * h[m,n] = \sum_k \sum_l x[k,l] + h[m-k, n-l]} \quad \textbf{(5.1)}$$

donde $x[m,n]$ es el campo receptivo de entrada al filtro, $h[m,n]$ es el filtro *kernel* que se aplica e $y[m,n]$ es la convolución que se obtiene a la salida del filtro.

La Figura **8**, muestra gráficamente el procedimiento de convolución, así como la variación entre las dimensiones de entrada y salida a dicha capa convolucional.

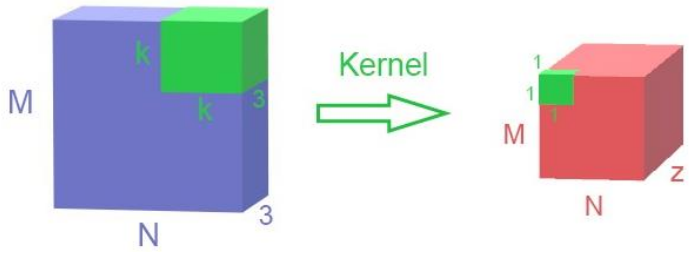


Figura 8: Representación del proceso de convolución sobre una imagen de entrada.

En las ocasiones donde las dimensiones de los filtros no coinciden con las de las imágenes, al aplicar la convolución, se reduce el tamaño de la imagen a la salida. Para controlar que esto no ocurra, se introduce un parámetro denominado *padding*. El *padding* define cómo se manipula el borde de una muestra. Una convolución *half (same) padding*, mantiene las dimensiones espaciales de salida igual que las de la entrada, mientras que convoluciones *unpadded* eliminan algunos bordes si el filtro es mayor que la unidad [26].

- ***Capa de activación***: Son capas no lineales situadas después de cada capa convolucional, dotando de no linealidad a los datos que se manejan. Es la capa donde se procesa la información, para ello se aplica una función de activación o umbral a la suma de los productos de las activaciones procedentes de las capas anteriores por sus pesos correspondientes [27]. Existen diversas funciones que se pueden emplear para la activación de mapas, entre las que se encuentran la sigmoide, la tangente hiperbólica, la lineal, leacky ReLU o Elu, entre otras, cuyas expresiones y rangos se pueden observar en la Tabla 2 [28].

- ***Capa de pooling***. Sirve para disminuir las dimensiones de entrada que le llegan, dejando inalterada la dimensión de profundidad. Cabe destacar que existen diversas operaciones para llevar a cabo el proceso de *pooling* [29]. La primera de ellas y más empleada es la operación *maxpooling*, encargada de tomar el nivel máximo de activación en el vecindario determinado por el campo receptivo. El tamaño típico de los filtros es 2x2, con un paso de 2. Suele utilizarse ya que lo que interesa conocer es la posición que ocupa una característica con respecto a las demás, y no tanto la localización precisa de dónde está situada en el espacio. En la Figura 9 se muestra el proceso de *maxpooling*. Con ello, se permite disminuir el coste computacional y ayuda a prevenir el sobreajuste.
  Otro método de *pooling* empleado es el denominado *average pooling*. Su funcionamiento es similar al de *maxpooling*, sin embargo esta vez la salida estará proporcionada por el valor medio de todos los componentes que integran el campo receptivo.

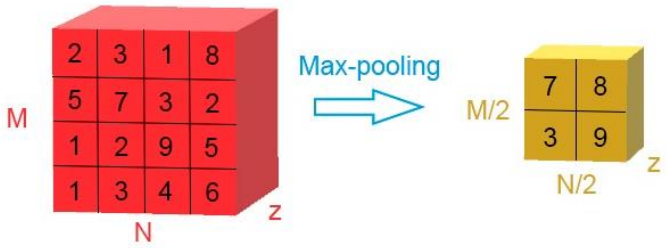


Figura 9: Representación del proceso de *Maxpooling* sobre un volumen de entrada.

- ***Capa completamente conectada (fully connected layer)***. Las capas descritas anteriormente conforman el *base model* de la arquitectura. Tras ellas, en el caso de redes neuronales destinadas a la clasificación, se deben conectar una serie de capas denominadas *fully connected* que conformarán el *top model*. De esta forma, las activaciones procedentes de la última capa del *base model* son estiradas a un único vector mediante una función denominada *flatten.* Esta modificación de volumen se puede observar en la Figura 10. De esta forma, las salidas tras aplicar la función *flatten* son llevadas a las capas *fully connected*, donde se encuentran las neuronas encargadas de llevar a cabo la clasificación.

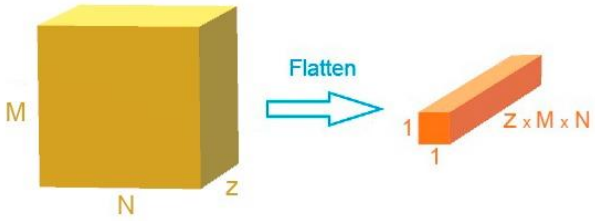


Figura 10: Representación del proceso de convolución unidimensional para la transformación del volumen de entrada a un único vector.

Al disponer de datos en capas completamente conectadas, se procede a aplicar una función de activación, comúnmente la función umbral *softmax*, que proporcionará la probabilidad de pertenecer a una determinada clase [29]. Se cumple que la suma total de cada una de las salidas es la unidad. Así se obtiene la probabilidad de distribución de las categorías. En la Figura 11, se muestra la Figura 11.

Figura 11: Representación de la activación *softmax* para obtener la clasificación final del modelo.

5.4 Proceso de aprendizaje de las redes neuronales convolucionales.

El proceso de entrenamiento de una red neuronal se divide en dos pasos fundamentales: propagación hacia delante, o *feedforward propagation*, y propagación hacia atrás o retropropagación, también conocido como *backpropagation* en su nomenclatura en inglés.

5.4.1 Propagación hacia delante

Es el primer paso a realizar, donde las imágenes de entrada son proporcionadas al sistema. Cada imagen recorre la arquitectura implementada, atravesando cada una de las capas de la red, y realizando todas las convoluciones planteadas. Una vez la imagen supera la capa *fully connected*, llegaría a la salida del sistema tras realizar la predicción con la función de activación *softmax*. Hasta este momento, la red no dispone de conocimiento. Es aquí donde se le proporciona la etiqueta a la red, tratándose por tanto de aprendizaje supervisado. Así, se realiza una comparativa entre la predicción estimada de la categoría a la que pertenece la imagen y su categoría real, para, en la siguiente fase, tratar de ajustar los pesos de todos los filtros de forma que las próximas predicciones se asemejen más a la realidad.

Este ajuste se ha de realizar atendiendo a la función de pérdidas que se empleda, tratando de minimizarla al máximo. Para ello, una de las técnicas más empleadas en clasificación es la entropía cruzada (5.2) que mide la diferencia que existe entre la distribución de probabilidad estimada y la real, de forma que cuanto menor sea, mayor será el parecido entre ellas. Parte de la definición de entropía propuesta por Shannon, haciendo referencia a este concepto como el mínimo tamaño de codificación que ha de tener un determinado mensaje de forma que no se pierda información. Así pues, la expresión analítica de la entropía cruzada es:

$$\boldsymbol{H(y,\hat{y}) = \sum_i y_i \log\left(\frac{1}{\hat{y}_i}\right) = -\sum_i y_i \log(\hat{y}_i)} \quad \textbf{(5.2)}$$

Siendo y_i la probabilidad de cada elemento, y $\hat{y}_i$ la probabilidad de la predicción.

5.4.2 Propagación hacia atrás o retropropagación:

En la fase de retropropagación, se ha de establecer una relación entre la aportación de cada peso de los filtros a la función de pérdidas [30]. Para ello, se implementa la derivada de la función de pérdidas respecto a los pesos de los filtros (5.3), para su posterior actualización.

$$\boldsymbol{\nabla J_i = \frac{\partial Error}{\partial w_i}} \quad \textbf{(5.3)}$$

La actualización de los pesos puede variar en función del optimizador y parámetros empleados en la fase de entrenamiento. Así, algunos de ellos se muestran en la Tabla 3.

Optimizadores	Ecuación auxiliar	Actualización de pesos
Actualización SGD		$W(t+1)=W(t)-\eta\nabla J$
Actualización SGD con Momentum	$V(t+1)=\gamma V(t)-\eta\nabla J(W(t))$	$W(t+1)=W(t)+V(t+1)$
Actualización SGD con Nesterov	$V(t+1)=\gamma V(t)-\eta\ \nabla J(W(t+\gamma V(t)))$	$W(t+1)=W(t)+V(t+1)$
Adagard	$c(t+1)=c(t)+\nabla J^2$	$W(t+1)=W(t)-\frac{\eta}{\sqrt{c(t+1)}+eps}\nabla J$ $eps\cong[10^{-4},10^{-8}]$

Tabla 3: Función de actualización de pesos atendiento al optimizador empleado

Donde η indica la tasa de aprendizaje o *learning rate* (lr). Así, una lr demasiado alta puede llevar a saltos tan grandes que la función de coste no converja hacia el mínimo global, mientras que tasas de aprendizaje pequeñas pueden ocasionar tiempos elevados de aprendizaje y mucho coste computacional.

Dichas fases de propagación hacia delante y retropropagación se han de realizar para un determinado número de épocas. Una particularidad a tener en cuenta es que cuando se trabaja con grandes bases de datos, el entrenamiento puede ser lento y costoso. En este contexto, se suelen emplear técnicas de agrupación para acelerar el proceso de entrenamiento. Para ello, lo habitual es generar pequeños lotes, denominados *batch*, de forma que una época finalizará cuando todos los *batches* hayan realizado las dos fases de propagación.

Además, durante la fase de entrenamiento, se trabajará con dos bases de datos, la de entrenamiento y la de validación, siendo esta última de dimensiones más reducidas. Los datos que componen ambas particiones son diferentes, de forma que pueda comprobar si el modelo adquiere conocimiento y capacidad de generalización.

Finalmente, para verificar que el modelo es capaz de predecir, se realiza una predicción final con las muestras que conforman la base de datos de test, desconocidas hasta el momento por la red. Así, se proporciona un informe con diferentes métricas en las que se puede comprobar si verdaderamente funciona o no el sistema.

Con todo lo visto hasta el momento, es posible construir una arquitectura neuronal para crear un modelo de clasificación. En el siguiente apartado se proponen diversas técnicas que pueden aportar grandes mejoras a los modelos predictivos.

5.5 Otras consideraciones a tener en cuenta

5.5.1 *Dropout*

Las redes neuronales profundas contienen numerosas capas no lineales que permiten encontrar relaciones entre sus entradas y salidas. Sin embargo, cuando se dispone de pocos datos, puede darse el caso de que el modelo aprenda una relación entre los parámetros de entrada y la salida, que no se corresponda realmente con la relación real existente entre ellos. En estos casos se dice que se produce sobreajuste u *overfitting*, cuando la respuesta de un modelo se ajusta en exceso a los datos específicos empleados para el entrenamiento del mismo, de forma que pierde su capacidad de generalizar a conjuntos de datos diferentes [31].

Para tratar de evitar que esto ocurra, Srivastava et al.[32] propusieron hacer uso de la técnica de *dropout*. Este mecanismo consiste en desconectar un determinado número de neuronas aleatoriamente, de forma que se reducen los grados de libertad de la red durante el proceso de entrenamiento. Así, el parámetro de dropout será un valor que indicará el porcentaje de neuronas a desconectar.

5.5.2 *Batch normalization*

Esta técnica es empleada para regularizar el modelo, la capa de entrada se normaliza ajustando y escalando las activaciones. Por ejemplo, cuando se disponen de características en un rango de 0 a 1 y otras en un rango de 1 a 1000, se normalizan para incrementar la velocidad de aprendizaje. Este mismo proceso se traslada a las capas ocultas de una red neuronal, denominándolo *batch normalization*, debido a que la distribución de las entradas de cada capa cambia durante el entrenamiento, reduciendo su velocidad [33].

Para aumentar la estabilidad de una red neuronal, este proceso normaliza la salida de una capa de activación anterior restando la media del *batch* y dividiéndola por su desviación estándar. Esta técnica añade dos parámetros entrenables a cada capa, de modo que la salida normalizada se multiplica por un parámetro de desviación estándar (*gamma*) y se añade un parámetro de media (*betha*). Con ellos, si debido al *batch normalization* los pesos de la siguiente capa dejan de ser óptimos, el optimizador puede cambiar estos dos parámetros en lugar de todos los pesos. En la Figura 12 se puede observar proceso gráficamente.

Gracias a esta técnica se pueden utilizar tasas de aprendizaje más elevadas, ya que garantiza que no haya ninguna activación con niveles desmesurados. Adicionalmente, puede reducir el *overfitting* debido a que tiene un ligero efecto de regularización.

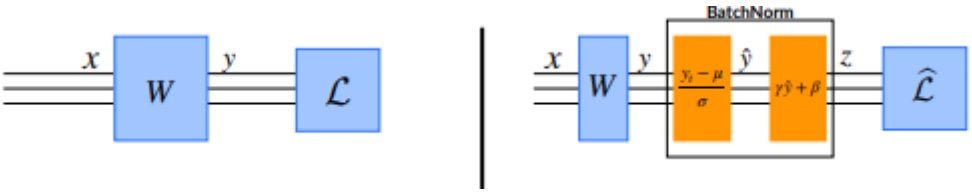


Figura 12. Proceso de *batch normalization*. Se puede observar el proceso de normalización, así como la adición de los dos nuevos parámetros [34].

5.5.3 Aumento sintético de datos

Esta técnica consiste en ampliar artificialmente la cantidad de datos disponible mediante transformaciones geométricas y de intensidad de color aplicadas a las imágenes reales. Así pues, se escogen pequeñas regiones de las imágenes reales y se les realiza diversas modificaciones, como pueden ser rotaciones, traslaciones, re-escalados y variaciones en el brillo, entre otras. De esta forma, el orden y amplitud de los pixeles se modifica radicalmente en cada una de las imágenes sintéticas creadas, por lo que serán percibidas como imágenes nuevas por la máquina. Así, se permite por una parte aumentar la base de datos con la que entrenar los modelos, y, por otra, evitar que se produzcan alteraciones como puede ser el sobreajuste.

5.5.4. Validación cruzada

Una vez se entrena el modelo, se necesita una garantía de la exactitud de las predicciones obtenidas. Para ello, es necesario validar el modelo. La validación cruzada es una de las técnicas utilizadas para evaluar la precisión y capacidad de generalización de un modelo así como un mecanismo de remuestreo que permite validar un modelo aunque se dispongan de datos limitados. Para llevar a cabo la validación cruzada, es necesario separar un conjunto de los datos que no se utilizarán para entrenar el modelo, sino para la posterior validación.

La validación cruzada de K iteraciones es una de las técnicas más populares resultando un modelo menos sesgado en comparación con otros métodos, dado que asegura que cada elemento del conjunto de datos original tenga la misma oportunidad de aparecer en el conjunto de entrenamiento y de validación [35]. El procedimiento es el que se muestra a continuación:

- Dividir el conjunto de datos de forma aleatoria en k subconjuntos (k normalmente suele tener un valor entre 5 y 10). Un valor alto de k resultará en un modelo menos sesgado, pero, puede conllevar también un sobreajuste.
- A continuación, se entrena el modelo utilizando los k-1 grupos y se valida con el restante. Se registran los resultados y errores.
- Se repite el proceso hasta que cada uno de los subconjuntos creados sirva como grupo de validación (es decir, repetir el proceso k veces). Una vez realizado todo el procedimiento, se obtiene el promedio de las k operaciones.

En la Figura 13, se muestra un ejemplo de *cross validation* para un *k-fold* = 3.

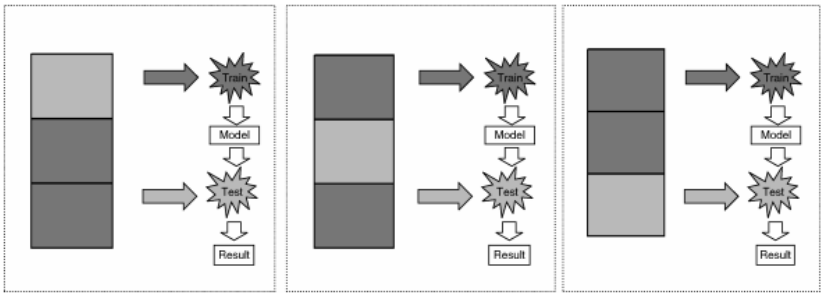


Figura 13. Ejemplo de *cross validation* con k=3 [35].

5.6 Transferencia de conocimiento y ajuste fino

Como ya se ha mencionado anteriormente, la gran limitación que presentan los modelos basados en redes neuronales convolucionales aplicados a imágenes, es la ausencia de grandes bases de datos etiquetados procedentes del problema al que se quiere dar solución. Además, en caso de disponer de un elevado número de imágenes, será necesario también tener acceso a potentes procesadores gráficos para operar con ellas. Para dar solución a este problema, surge la transferencia de conocimiento y el ajuste fino [36] .

Comúnmente conocida como *transfer learning*, la transferencia de conocimiento consiste en re-entrenar arquitecturas que anteriormente ya han sido entrenadas con grandes bases de datos etiquetados. Esto permite hacer uso de filtros con pesos ya ajustados y optimizados para las imágenes del modelo de origen, de modo que ya disponen de conocimiento. Estos pesos son proporcionados al nuevo modelo para que, tras una fase de entrenamiento, se ajusten de manera fina a las nuevas imágenes que se desea clasificar.

Para ello, se inicializan los pesos de la primera capa del *base model* con los pesos procedentes de la transferencia de conocimiento, y se modifican las capas del *top model*, ajustando las neuronas de la capa *fully connected* para adaptarlas al número de categorías en las que clasificará el nuevo modelo diseñado.

Cuando se modifica alguna de las capas de la arquitectura del modelo original, se dice que se está realizando *fine tuning*, o ajuste fino. Con ello, lo que se pretende es no utilizar los pesos finales del modelo origen directamente en el nuevo clasificador, sino únicamente los obtenidos en algunas de las capas en concreto y re-entrenar el resto de capas de la arquitectura con las nuevas imágenes. Así pues, si solo se modifica la última capa de la arquitectura, se trataría de un ajuste poco profundo o *Shallow tuning*, mientras que, si se re-entrenan más capas, se estaría realizando ajuste profundo o *Deep tuning*. Por tanto, el ajuste será más profundo cuantas más capas se re-entrenen.

Para poner en práctica esta técnica de aprendizaje, cualquier arquitectura que haya sido entrenada con una base de datos consistente puede ser empleada. No obstante, comúnmente se utilizan las arquitecturas convolucionales resultantes del concurso anual de ImageNet, ofrecidas al público gratuitamente. En el "ImageNet Large Scale Visual Recognition Challenge" (ILSVRC), cada año, los participantes proponen nuevas arquitecturas con el fin de mejorar la precisión alcanzada de los modelos previos, reduciendo de igual modo el error de clasificación. Concretamente, en el ámbito de clasificación que nos ocupa, la base de datos de ImageNet ofrecida por *Google*, está compuesta por un total de 1.35 millones de imágenes, de las cuales 100.000 son destinadas para testeo y 50.000 para validación; todas ellas pertenecientes a 1000 categorías. Se trata, por tanto, de modelos robustos con amplio conocimiento en sus pesos[37].

De esta forma, en 2012 surge AlexNet, propuesta por Krizhevsky et al.[38] , con una profundidad de 8 capas. De todas ellas, cinco son convolucionales activadas con la función ReLU, y las tres capas restantes son completamente conectadas, que, junto con tres capas de *max-pooling*, dan como resultado la clasificación final. Con ella, obtuvieron un error de clasificación un 10% inferior a la que alcanzó el segundo puesto el año anterior.

Ya en 2014, tres grandes arquitecturas fueron propuestas, la VGG16 y VGG19, así como la GoogleNet. Las dos primeras, parten de la idea de que la precisión de clasificación puede ser mejorada dotando de profundidad a la red, mientras que la última se corresponde con la arquitectura Inception, lo que le permitió a Szegedy et al.[39], encabezar la tabla de clasificación de 2014. Ambas arquitecturas consiguieron reducir el error que obtuvo AlexNet en 2012.

Justo un año después, en el ILSVRC 2015, la competición fue liderada por Microsoft's Residual Netwok, ResNet con un indicador de error considerablemente inferior a las ganadoras de las competiciones anteriores [40].

Como ellos, otros muchos autores tratan de encontrar arquitecturas con una base sólida de conocimiento que favorezca al diseño de nuevas aplicaciones en las que no se dispone de datos en abundancia. A continuación, se procederá a explicar las grandes arquitecturas que han destacado en el concurso ImageNet, y que después serán empleadas para la implementación de del modelo de clasificación distintiva entre ojos sano y ojos glaucomatosos, idea principal que impulsa al desarrollo de este trabajo.

5.6.1 VGG16 y VGG19

El Grupo de Geometría Visual de la Universidad de Oxford introdujo el modelo VGG, de la mano de K.Simonyan, y A. Zisserman. La gran contribución que aportaron dichos autores, fue aumentar la profundidad de la red hasta alcanzar las 16-19 capas de profundidad, dando lugar a las arquitecturas VGG16 y VGG19, respectivamente. Para ello, hicieron uso de una arquitectura con filtros convolucionales muy pequeños, concretamente de tamaño 3x3 [41].

Como entradas, emplearon imágenes de tamaño 224x224 en escala RGB que, tras una fase previa de preprocesado, se introducen en un conjunto de capas convolucionales con filtros de campo receptivo 3x3 y paso 1, lo que indica que la convolución se aplica sobre cada pixel. Así pues, la estructura que presentaron consistía en un total de trece capas convolucionales en el caso de la VGG16 y dieciséis en la VGG19, con los filtros 3x3 apilados uno encima de otro para dotar de profundidad a la red, seguidos de la función ReLU para la activación de los pesos. Además, entre determinadas capas convolucionales, utilizan una capa de *maxpooling* en ventanas 2x2 con paso 2; en total, se pueden encontrar cinco capas de este tipo. Finalmente, le sigue un conjunto de tres capas *fully connected*, donde la última presenta 1000 neuronas, correspondientes a las 1000 categorías de los datos de ImageNet. Luego le procede una capa de *softmax*, donde se obtienen las diferentes probabilidades.

En total los autores presentaron 6 arquitecturas diferentes, siendo las dos señaladas en la Tabla 4, las que poseen mayor profundidad, y las que mejor precisión obtuvieron. Todo ello, les permitió alcanzar el primer y segundo puesto en la sección de clasificación de imágenes.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Tabla 4: Detalle de las diferentes capas de las arquitecturas VGGx. La arquitectura D se corresponde con la VGG16, mientras que la E identifica a la VGG19 [41].

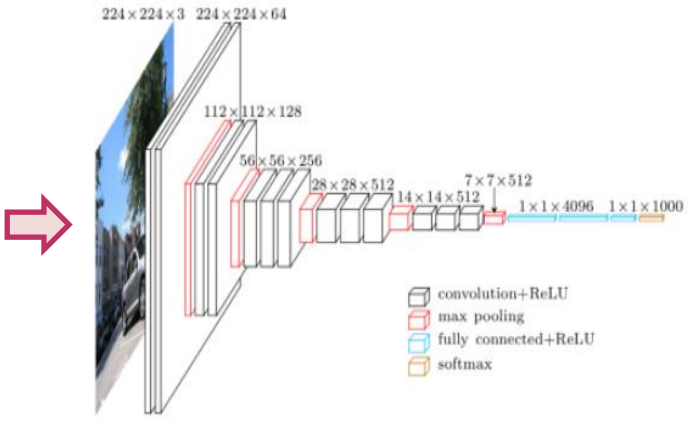


Figura 14: Arquitectura VGG16

5.6.2 Inception V3

En el ILSVRC de 2014, el primer puesto en la categoría de detección de objetos, fue otorgado a la arquitectura Inception de GoogleNet [39]. Esta red se basa en la utilización de filtros convolucionales cuadrados de diferentes tamaños, que finalmente son agregados para dar un resultado final.

Concretamente, las dimensiones de los campos receptivos son 1x1, 3x3 y 5x5, para la extracción de características. Además, se aplica una convolución adicional de tamaño 1x1 a las imágenes de entrada, de forma que queden recogidas bajo un único canal. Seguidamente a las capas convolucionales, se aplica una capa de *maxpooling*. En la Figura 15, se muestra la arquitectura de un bloque de Inception.

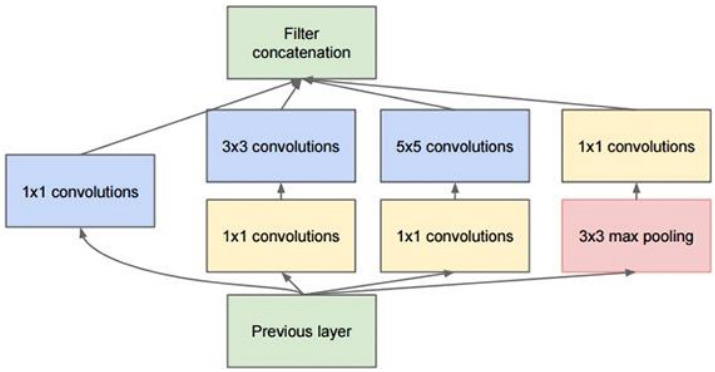


Figura 15: Módulo de Inception [42].

Tras la creación de esta arquitectura, nuevas redes fueron surgiendo, tomando como base las estructura de Inception. Se basan, por tanto, en la modificación de las dimensiones de los filtros para tratar de disminuir el error a la vez que se mejora el coste computacional. De esta forma, surge Inception V3, que ocupó el primer puesto en la clasificación del ILSVRC 2015, con un error del 4.2% en la sección de clasificación, frente al 7.89% de GoogLeNet y el 6.8% de VGG16, en el ranking Top-5 error en testeo [42].

Así pues, Inception V3, se trata de una arquitectura compuesta por 42 capas. Su estrategia consiste en remplazar cada filtro de tamaño 5x5 empleados en los modelos InceptionVx anteriores por dos filtros 3x3 en serie, dando lugar a una reducción total de parámetros del 28 % (un campo receptivo de 5x5 da lugar a 25 parámetros, mientras que uno de 3x3 genera 9 parámetros). También se puede reducir los parámetros empleando convoluciones asimétricas. Una convolución 3x3 equivale a una 3x1 seguida de otra 1x3, disminuyendo en este caso los parámetros un 33% (véase un ejemplo en la Figura 16). Al reducir el número de parámetros, la arquitectura es menos propensa a presentar sobreajuste. Además, añadieron un clasificador auxiliar, que regularizaron aplicándole *batch normalization*.

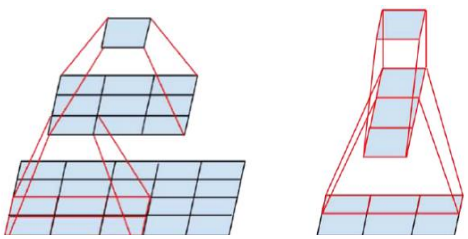


Figura 16: Representación gráfica de las equivalencias de dimensiones en las convoluciones. A la izquierda, se aprecia como dos convoluciones 3x3 dan lugar a las mismas dimensiones que una de 5x5. A la derecha, una convolución 3x1 seguida de una 1x3 da como resultado una de 3x3 [42].

5.6.3 Xception

En el año 2016, Chollet el al. [43] propusieron una nueva arquitectura, que surge como extensión de la Inception de GoogLeNet. De hecho, tal y como indican sus creadores, el nombre "Xception" procede de "*Exteme Inception*".

La idea subyacente en la creación de esta red, parte de que las correlaciones existentes entre las dimensiones espaciales altura y anchura tras una convolución, con respecto a la profundidad, se encuentran lo suficientemente desacopladas entre ellas, por lo que se procede a activarlas por separado (en InceptionV1 se mapeaban todas en la misma capa, obsérvese Figura 155). En la Figura 17, que se muestra a continuación, se observa un esquema de cómo se realizaría el mapeo individualmente.

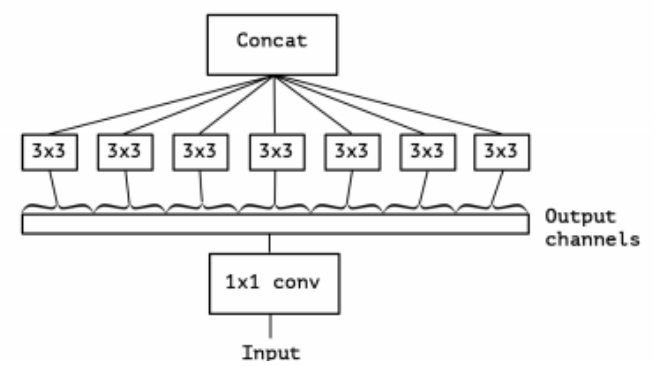


Figura 17: Módulo de Xception, versión extrema de un módulo Inception con una convolución espacial por canal[43].

De esta forma, la arquitectura Xception, está conformada por 36 capas convolucionales, con filtros de tamaño fijo 3x3 y paso 2, divididas en un total de 14 bloques.

Así, Xception igualó el número de parámetros obtenidos con la arquitectura Inception V3, si bien los resultados de precisión y error de Xception superan a los de su antecesor.

5.6.4 ResNet

Propuesta por He et al. en el ILSVRC 2015, la red residual ResNet quedó en primera posición para las categorías de clasificación y detección de objetos [40]. Se trata de una red menos compleja que la VGG16 y VGG19.

Su propuesta surge de la necesidad de solucionar el problema de capacidad y saturación dependiente del aumento de profundidad de las arquitecturas. Lo que propusieron, fue hacer uso de bloques residuales convolucionales cuya principal utilidad que aporta es que la salida de cada bloque se realimenta con las entradas al mismo antes de dar comienzo al siguiente bloque. En cada uno de estos bloques, utilizaron campos receptivos de 3x3 y paso dos. Además, entre cada convolución y la activación siguiente, incorporan *batch normalization* como método de regularización.

En la Figura 18 que se muestra a continuación, se puede observar la estructura de un bloque residual. El de la izquierda se corresponde con el modelo ResNet 34, compuesto por dos capas convolucionales. Otras propuestas de esta arquitectura son ResNet 50/101/152, en los cuales el bloque contiene tres capas. Es necesario destacar que se trata de una red menos compleja que la VGG16 y la VGG19, ya que empleando *global average pooling* y capas *fully connected* en su estructura requiere de menos capacidad de almacenamiento, a pesar de ser una red más profunda. Con todo ello, consiguió batir records con una tasa de error de tan solo el 3.57% en el top-5 error en testeo.

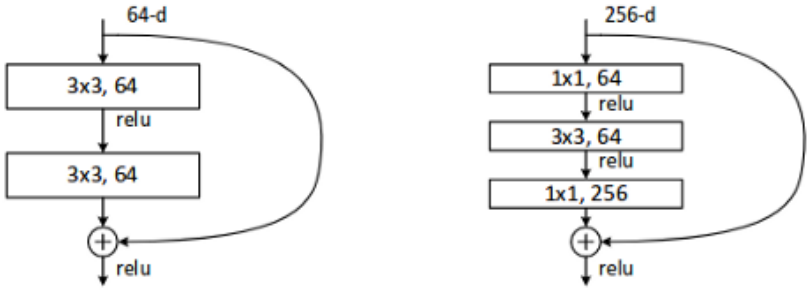


Figura 18: Arquitectura de bloques residuales de ResNet

5.7 *Machine Learning* en el diagnóstico automático del glaucoma a partir de imágenes OCT.

En las últimas dos décadas, se han llevado a cabo gran variedad de investigaciones con el objetivo de ayudar a médicos y especialistas oftalmólogos a la detección y diagnóstico de glaucoma. De esta forma, la mayoría de artículos existentes hasta el momento que destinan sus estudios a la identificación de glaucoma teniendo como base parámetros obtenidos a partir de la herramienta OCT, utilizan técnicas tradicionales de *machine learning*, como pueden ser *Random Forest* [44], las máquinas de soporte vectorial [45] y *K-Nearest Neighbor* [46]. Así pues, Kim et al. [47] propusieron diversos modelos para la detección de dicha patología, empleando estos tres algoritmos de clasificación, haciendo uso para ello de los espesores de la capa de fibras nerviosas de la retina obtenidos en la OCT y el campo visual de los pacientes.

Otros autores como Silva et al. [48] y Barella et al. [49], combinan parámetros de espesores del nervio óptico y de la capa RNFL obtenidos a partir de OCT, con otro tipo de datos demográficos (sexo, edad, raza) y pruebas médicas que determinan la presión intraocular, la agudeza visual, el grosor real de la retina del paciente, entre otros, que serán la entrada a los clasificadores.

Con respecto a las técnicas de clasificación basadas en aprendizaje profundo, suelen emplearse redes neuronales convolucionales. En la literatura, encontramos autores que realizan sus investigaciones con imágenes de fondo de ojo [50][51], mientras que otros hacen uso de esta técnica para segmentar las capas de la retina y obtener los espesores [52], que más tarde utilizarán para detectar la presencia o no de glaucoma.

En [53], Muhammad et al. presentaron un modelo de predicción híbrido basado en aprendizaje profundo en el que combinaron un extractor de características implementado con redes neuronales convolucionales, al que incorporaron un clasificador *random forest*. Los datos de entrada son imágenes que representan el mapa de probabilidad de la capa retiniana, obtenido a partir de las imágenes cirmumpapilares de la oct y los espesores de las capas de la retina. Como método de extracción de características, utilizaron la técnica de *transfer learning* a partir de la arquitectura AlexNet y *data augmentation* con volteo vertical.

En el caso del presente trabajo fin de grado, se van a proponer una serie de modelos para la detección automática de glaucoma basadas en aprendizaje profundo únicamente, con el empleo de imágenes circumpapilares. Este trabajo se enmarca dentro del proyecto europeo GALAHAD [H2020-ICT-2016-2017, 732613] [1]. La idea que persigue dicho proyecto, es el diseño de un dispositivo OCT de bajo coste para incorporar en los centros de atención primaria, de forma que sirva como guía para determinar si un paciente presenta o no signos de glaucoma, y en caso afirmativo, destinarle al médico especialista.

Capítulo 6. Modelos predictivos basados en *Deep learning* para la identificación del glaucoma.

Llegados a este punto, se va a proceder a describir los modelos propuestos para la clasificación automática de glaucoma. Para ello, dos arquitecturas han sido la guía de este problema, una de ellas diseñada y entrenada desde cero y otra haciendo uso de transferencia de conocimiento. Así pues, se va a describir la composición estructural de las mismas, así como las modificaciones más importantes llevadas a cabo para la determinación del modelo final. Posteriormente, se tratará de incorporar datos demográficos además de las imágenes circumpapilares propiamente dichas, para para determinar si estos suponen algún tipo de repercusión en la clasificación.

6.1 Material

6.1.1 Base de datos

Una tarea esencial a realizar previamente a la creación de modelos consiste en conformar una base de datos amplia y consistente, ya que los resultados obtenidos dependerán directamente de esta.

En este caso, los datos que emplearemos proceden de la clínica oftalmológica Oftalvist, quienes proporcionaron al *Computer Vision and Behaviour Analysis Lab* imágenes de pacientes sanos e imágenes de pacientes que padecen glaucoma. El formato de origen de dichas imágenes es propietario, concretamente los ficheros son de extensión ".e2e", lo que requiere una tarea inicial de descompresión y conversión hasta alcanzar el formato deseado, en este caso, imágenes con extensión ".tif". Para ello, se ha hecho uso del software Heidelberg, especializado en descomprimir y tratar imágenes de OCT. Así, se toma el archivo de cada paciente y se introduce en el programa. Desde dicho *software* se pueden llevar a cabo diversas tareas, entre las que se encuentran las de visualización y segmentación de capas del nervio óptico.

En nuestro caso, se ha llevado a cabo una segmentación mediante Heidelberg debido a dos factores clave. El primero de ellos es que atendiendo a la definición de glaucoma, cuando este aparece se produce un aumento de la excavación de las células ganglionares de la retina, con la consiguiente reducción del espesor del nervio óptico. El segundo de ellos es que Heidelberg dispone de una amplia base de datos de pacientes sanos y patológicos, así como del espesor de sus capas retinianas. Por este motivo, para un nuevo paciente, se obtiene la curva que representa las capas superior e inferior de su retina, se obtiene el espesor medio y Heidelberg lo compara automáticamente con pacientes de la misma edad y sexo.

De esta forma, al realizar la segmentación automática en el software, este proporciona un gráfico sectorial en el que indica si el paciente se encuentra dentro de los límites que corresponden a los pacientes sanos, o si por el contrario se encuentra fuera de los límites, lo que conllevaría a una posible patología oftalmológica.

En la Figura 19, se pueden observar los gráficos proporcionados por el *software* Heidelberg tras la entrada nuevas de imágenes circumpapilares del nervio óptico.

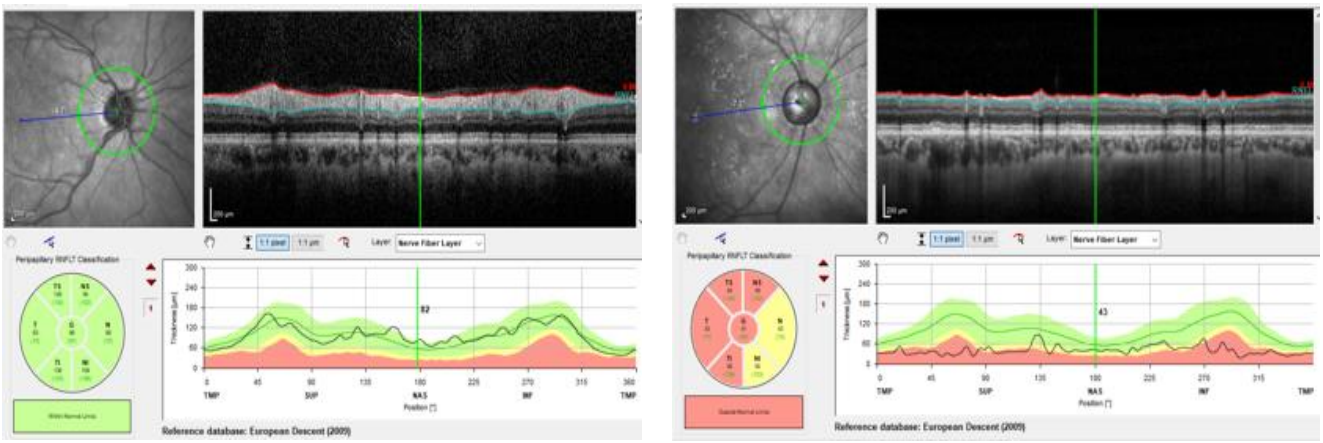


Figura 19: Gráfico sectorial y espesor medio de la capa de la retina representado por Heidelberg a partir de la imagen circumpapilar de cada paciente. A la izquierda, un ojo sano, a la derecha, uno glaucomatoso.

Por tanto, se ha realizado este procedimiento para cada uno de los pacientes proporcionados por Oftalvist, de modo que los pacientes sanos son los entregados por la clínica, que a su vez el programa nos indica que efectivamente lo son, mientras que los pacientes con glaucoma son los que la clínica nos entrega como muestras patológicas y que además el programa nos indica que el espesor de su retina se encuentra fuera de los límites normales. A continuación, se exportan los datos recibidos a archivos con extensión ".xml", con los que, tras realizar un procesado en Matlab dan lugar a la imagen en formato ".tif".

Como resultado de este proceso, se obtiene una base de datos de 246 imágenes en escala de grises, de las cuales 162 son muestras sanas, mientras que las 84 restantes pertenecen a pacientes con glaucoma. A priori, esto proporciona un claro desbalanceo entre clases, por lo que habrá que tratarlo convenientemente durante el proceso de aprendizaje de nuestro modelo.

Una vez disponemos de los datos necesarios, se procede a realizar la partición de los mismos. Como se ha mencionado en la Sección 5.4, se requiere dividir los datos en tres bloques, cada uno de los cuales representará la base de datos de entrenamiento, de validación o de test.
Además, dado que poseemos pocos datos y el problema a resolver requiere de gran precisión, y poseer la habilidad de generalización, se ha decidido realizar validación cruzada. Por este motivo, crearemos cinco particiones de datos, de forma que, en cada una de ellas, destinaremos el 20% de los datos disponibles para testeo, mientras que el 80% restante pertenecerá un porcentaje a validación y otro a entrenamiento, en una proporción 20%-80%, respectivamente. Al realizar esta división de forma automática, el número de imágenes destinadas a las diferentes bases de datos, así como a cada una de las cinco particiones se muestra en la Tabla 5, donde S se corresponde a pacientes sin glaucoma y G indica el caso contrario.

	Carpeta 1		Carpeta 2		Carpeta 3		Carpeta 4		Carpeta 5	
	S	G	S	G	S	G	S	G	S	G
Entrenamiento	103	53	103	53	104	53	104	53	104	54
Validación	26	14	26	14	26	14	26	14	26	14
Test	33	17	33	17	32	17	32	17	32	16

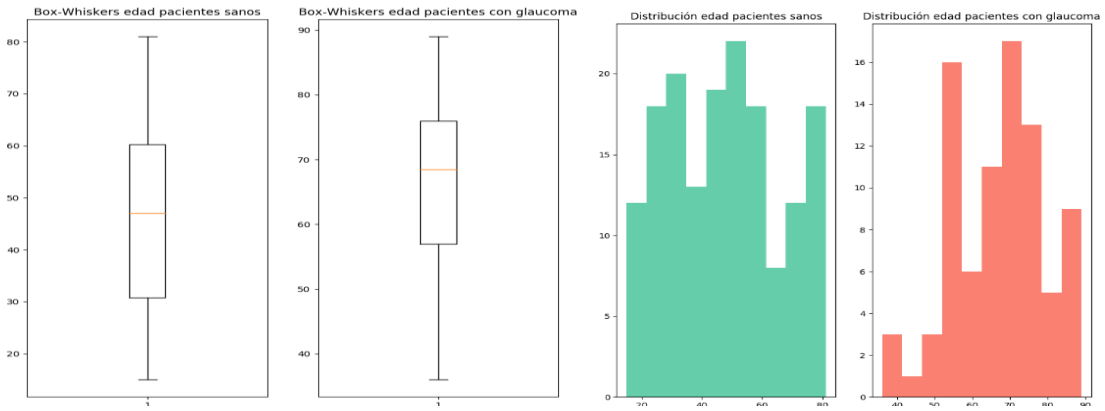
Tabla 5: Partición de los datos destinados a entrenamiento, validación y testeo, para cada una de las cinco particiones de validación cruzada.

Por otra parte, es necesario realizar una normalización de los datos. Las imágenes están compuestas por píxeles, cuyo color se define por la amplitud de los mismos, que se encuentra en el rango [0-255], donde 0 se corresponde con el color negro y 255 con la máxima intensidad, es decir, blanco. Para que todos los píxeles sean igual de relevantes para nuestra red de clasificación, debemos normalizarlos en el rango de [0,1] por lo que dividiremos cada pixel entre 255. Además, tras realizar diferentes simulaciones en los modelos planteados se decidió redimensionar el tamaño de las imágenes, pasando de unas dimensiones 496x768 a imágenes cuadradas de 256x256 ya que estas proporcionaron mejores resultados.

Por otra parte, las etiquetas glaucoma y sano serán identificadas como ‘0’ o ‘1’, respectivamente. En el momento de entrenar el modelo y en el de obtener las predicciones, será necesario convertir estos datos a formato *one hot encoding*.

6.1.2 Análisis de variables demográficas

Además de las imágenes circumpapilares, disponemos de dos variables adicionales que nos indican la edad y el sexo de cada paciente. De esta forma, podemos analizar estadísticamente si existe, a priori, una relación lineal entre dichos datos y que el paciente presente o no glaucoma. Así, analizando parámetros de posición y dispersión de la variable edad, no se encuentra un claro patrón que permita discernir entre pacientes sanos y patológicos, como era de esperar. Sin embargo, se puede observar que la media de edad de pacientes sanos se encuentra en torno a 47 años, mientras que la de pacientes patológicos en 67, si bien la moda es de 42 en el primer caso y 72 en este último. Además, el rango de edad de ambas categorías es bastante amplio, superior a 50 años en ambos casos, tal y como se aprecia en el diagrama *box-whiskers (*ver Gráfica 1 *y* Tabla 6*)*

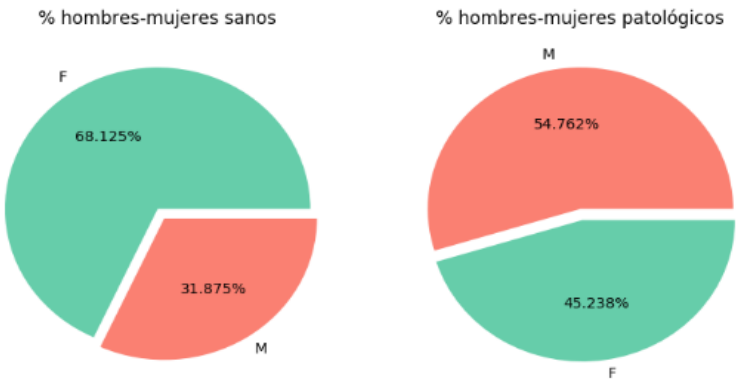


Gráfica 1: Diagrama box-whiskers e histograma de la variable edad en pacientes sanos y partológicos.

	Sano	Glaucoma
Promedio	47.38	67.1
Mediana	47	68.5
Moda	42	72
Desviación Estándar	18.25	12.31
Varianza	333.1	151.58
Coeficiente de Variación	0.38	0.18
Mínimo	15	36
Máximo	81	89
Rango	66	53

Tabla 6: Parámetros de posición y dispersión de la variable edad.

En cuanto a la variable categórica que indica el género del paciente, el 68.125% de pacientes sanos son mujeres frente al 32.875% de varones. Por el contrario, en cuanto a pacientes con glaucoma se refiere, el 54.762% son hombres frente al 45,218% de mujeres. (Ver Gráfica 2).



Gráfica 2: Proporción de hombres y mujeres en cada categoría, donde 'M' representa al género masculino y 'F' al género femenino.

De esta forma no se encuentra ningún patrón que identifique el perfil de pacientes con glaucoma ateniendo únicamente a variables demográficas. No obstante, se observa que los pacientes con glaucoma presentan una edad ligeramente superior a los pacientes sin dicha patología (teniendo en cuenta que hay gran varianza entre las muestras), al igual que en frecuencia relativa, el género masculino que conforma nuestra base de datos presenta más casos de glaucoma que las mujeres.

6.1.3 Entorno de programación

El proyecto ha sido llevado a cabo en un equipo compuesto por un procesador intel® Core™ i7-4500U @1.8 GHz con *Turbo Boost* hasta 3 GHz y una memoria RAM de 8 GB de capacidad. Además, el equipo cuenta con una tarjeta gráfica AMD Radeon R5 M200 Series. En lo que respecta al sistema operativo, se trata de un Windows 10 de 64 bits.

Para el desarrollo e implementación de las diferentes propuestas de modelos, se ha hecho uso del lenguaje de programación Python3.5, para lo que se ha utilizado PyCharm 2019.1.2x64 como entorno de desarrollo integrado. En lo que al *software* respecta, se ha empleado como *framework* de desarrollo *Keras* que dispone de implementaciones de alto nivel para la construcción de redes neuronales, tales como les redes neuronales recurrentes y redes neuronales convolucionales [54], con *tensorflow* como *backend*. Además, para la etapa de entrenamiento de las redes neuronales convolucionales basadas en aprendizaje profundo se ha hecho uso de la librería NVIDIA Cuda® Deep Neural Network.

Otra de las librerías necesarias para la implementación de estos modelos son las librerías OPENCV con soporte CUDA, al igual que otros módulos del *framework sklearn* que nos permitirá obtener métricas de evaluación para los modelos generados [55].

Cabe añadir que debido a la complejidad computacional de las técnicas de aprendizaje profundo empleadas, se ha hecho uso de un servidor de computación de altas prestaciones compuestos por un procesador Intel i7 @4.20GHz, 32GB de RAM y tarjetas gráficas NVIDIA Titan V, que se encuentran disponibles en el CVBLab. Adicionalmente disponen de un servidor (Synology DS416) en el que se almacenan los datos referentes al proyecto, imágenes en este caso.

6.2 Modelos propuestos

En esta sección, se va a proceder a exponer los diferentes modelos que se han planteado, definiendo la arquitectura de cada uno de ellos, así como las diferentes modificaciones realizadas para tratar de mejorar la precisión obtenida en la clasificación.

Así pues, se va a crear inicialmente un modelo *from scratch*, y a continuación otro que haga uso de transferencia de conocimiento, para poder comparar resultados.

En la creación modelos, se ha de llevar a cabo una serie de tareas. Lo primero de todo será preprocesar los datos, como ya ha sido comentado, de forma que obtengamos tres bases de datos que son entrenamiento, test y validación, todas ellos normalizados y con las dimensiones deseadas.

Posteriormente, se procede a diseñar la arquitectura de la red, en la que se definen las capas que han de tener, tanto cuantitativa como cualitativamente aportando una serie de parámetros como es el tamaño del *kernel*, el *dropout* en caso de que lo hubiera, el *padding*, *stride*, etc.

Una vez definida la arquitectura y las entradas de la misma, se compila el modelo, y, tras ello comienza la fase de entrenamiento. A partir de este momento, darán comienzo las fases de propagación hacia delante y retropopagación para cada una de las épocas. Una vez finaliza el entrenamiento, se procede a la predicción con la base de datos de test. Para concluir, se calculan una serie de métricas con el objetivo de evaluar el modelo. Este proceso queda resumido de forma esquemática en la Figura 20.

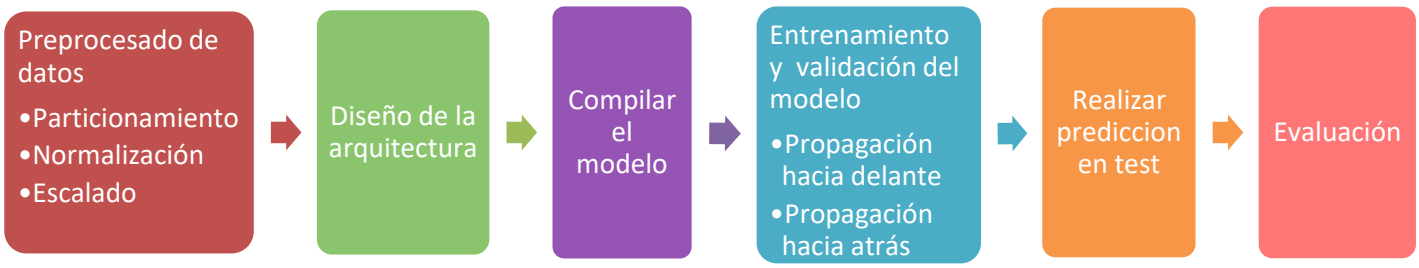


Figura 20: Proceso a llevar a cabo en la creación de un modelo de predicción con inteligencia artificial.

6.2.1 Creación de un modelo neuronal convolucional *from scratch*

La creación de un modelo neuronal convolucional *from scratch*, significa construir un modelo desde cero, es decir confeccionar nuestra propia arquitectura (ver Figura 21).

Para ello, se ha decidido hacer uso de dos bloques convolucionales. Cada uno de ellos, está compuesto por dos capas convolucionales, con campos receptivos 3x3 y paso 1, tras los que se aplica la activación ReLU y *padding same*. La diferencia existente entre los dos bloques es el número de filtros a emplear, en el primero de ellos se hará uso de 32 filtros, mientras que el segundo estará compuesto de 64 *kernels*.

Entre cada capa convolucional, se ha decidido introducir *batch normalización* para tratar de estabilizar la red neuronal. Tras cada uno de estos dos bloques, se aplica una capa de *pooling*. Concretamente la técnica empleada será *maxpooling* de tamaño 2x2 con paso 1, por lo que las dimensiones de entrada a esta capa quedan reducidas a la mitad a la salida.

Entre el primer y el segundo bloque se decide introducir *drop-out* con el fin de evitar el sobreajuste, en la proporción de 0.25, es decir, el 25% de las neuronas quedarán desconectadas aleatoriamente.

Tras la salida de la segunda capa de *pooling*, nos adentramos a las capas *fully connected*, donde se procederá a la clasificación. Para ello, se ha hecho uso de una capa *flatten*, que convierte el volumen procedente de la salida de la capa de *pooling* a un único vector. Tras él, conectamos dos capas *fully connected*, una de 100 neuronas a la que se aplicará la activación *ReLU*, y otra de 2, cada una de las cuales se corresponde con una categoría en las que buscamos clasificar, es decir, sano o patológico. Seguidamente, se aplica la activación *softmax* que proporcionará la probabilidad de pertenecer a cada clase. La composición detallada de cada una de estas capas se puede apreciar en la Tabla 7.

	Capas	Neuronas	Tamaño	Otros parámetros	Dimensiones de salida
1	Convolucional	32	[3,3]	Paso=[1,1], Activación='ReLU' Padding='same', Batch-Normalization	(256,256,32)
2	Convolucional	32	[3,3]	Paso=[1,1], Activación='ReLU' Padding='same', Batch-Normalization	(256,256,32)
3	Max-Pooling	-	[2,2]	Paso=[2,2], Dropout=0.25	(128,128,32)
4	Convolucional	64	[3,3]	Paso=[1,1], Activación='ReLU' Padding='same', Batch-Normalization	(128,128,64)
5	Convolucional	64	[3,3]	Paso=[1,1], Activación='ReLU' Padding='same', Batch-Normalization	(128,128,64)
6	Max-Pooling	-	[2,2]	Paso=[2,2], Dropout=0.25	(64,64,64)
7	Fully-Connected	100	-	Activación='ReLU', Dropout=0.5, Batch-Normalization	(,100)
8	Fully-Connected	2	-	Activación='Softmax' -	(,2)

Tabla 7: Definición de la arquitectura *from scratch*: capas y parámetros.

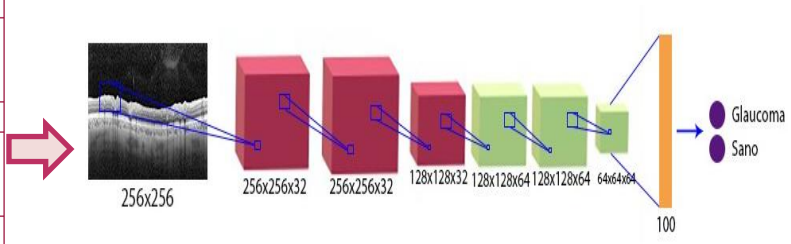


Figura 21: Arquitectura *from scratch*

El entrenamiento del modelo se lleva a cabo durante 1000[5] épocas, empleando el optimizador SGD con una tasa de aprendizaje de $\eta = 0.001$, pues es la que mejor se ajusta al modelo después de realizar una optimización empírica. Cabe destacar que para el proceso de entrenamiento se emplea también un tamaño de *batch* de 32, junto con la función de pérdidas *binary-crossentropy*, para lo que se hará uso de la expresión (5.2).

Por otro lado, tal y como comentamos en la Sección 6.1.1, existe un desbalanceo entre clases, ya que, del total de casos, el 34.15% presentan glaucoma, frente a un 65.85% de pacientes sanos. Para compensarlo, incorporamos un índice como parámetro de entrenamiento que asigna una ponderación diferente a cada una de estas clases, cuyo valor será tenido en cuenta en el momento de evaluar la función de pérdidas. Dicho de otro modo, servirá para que el modelo presente más atención a las muestras de la clase con menor representación.

Tras compilar y entrenar el modelo, se representa la evolución de las pérdidas y precisión que se produce tanto en entrenamiento como en validación conforme avanza en número de épocas de entrenamiento. En este momento, se aprecia la existencia de signos de *overfitting*.

Por este motivo, se decide aplicar *data augmentation*. De esta forma, se amplía el número de datos disponibles para una misma etiqueta, hecho que favorecerá a la reducción del sobreajuste. Por lo tanto, crearemos imágenes sintéticas incorporando inversiones horizontales, aplicando zoom en un ratio de 0.2, con un ángulo de corte de 0.2. En total, obtendremos 20 muestras sintéticas por cada imagen original. Re-entrenado la arquitectura anteriormente diseñada con el aumentado de datos, la precisión en clasificación mejora, si bien el sobreajuste, a pesar de reducirse, no desaparece.

Este hecho nos impulsa a considerar nuevas alternativas. Una de ellas consiste en incorporar datos adicionales al entrenamiento, como son la edad y el sexo de los pacientes; mientras que la segunda alternativa es hacer uso de la trasferencia de conocimiento. Ambas propuestas han sido planteadas, y serán expuestas en los apartados siguientes.

[5] En el momento de iniciar el entrenamiento del modelo, se indica un criterio de parada, de forma que en el momento en que las pérdidas dejen de disminuir durante 80 épocas consecutivas (hecho que indicaría que la red no aprende más), se finalizará el entrenamiento, independientemente del número de épocas restantes hasta alcanzar las 1000 definidas.

6.2.2 Combinación de una red neuronal convolucional *from scratch* y un perceptrón multicapa

Para intentar mejorar la precisión del modelo *from scratch*, se decide realizar una modificación en la arquitectura, de modo que se combine la información extraída de las imágenes, junto con datos de los pacientes, tales como la edad y el sexo de los mismos. En la generación de glaucoma, el desgaste de las capas del nervio óptico debidos a la edad juega un papel importante, del mismo modo que los especialistas afirman que pueden existir diferencias entre géneros. Basándonos en esta información, planteamos la hipótesis de que este tipo de datos puede contribuir a la correcta identificación de glaucoma.

De esta forma, diseñamos una segunda arquitectura en dos etapas (ver Figura 22). La primera de ellas consiste en entrenar la red neuronal convolucional del mismo modo que lo hicimos en la sección anterior. Hay que tener en cuenta que en este caso no aplicaremos la técnica de generación sintética de datos, ya que las imágenes han de corresponderse con un determinado paciente, con su respectiva edad y sexo. El hecho de aplicar modificaciones como rotaciones, giros o cambios en la intensidad de color para la creación de muestras sintéticas no modifican la categoría a la que pertenece la imagen, pero sí que pueden afectar de tal forma que no se relacionen directamente con la edad del paciente que se apreciaba en la imagen original.

El entrenamiento de la red convolucional con imágenes ejercerá de extractor de características, de forma que, al concluir el entrenamiento se extraerá la información existente en la capa *fully connected*, previamente a la clasificación.

A continuación, se concatenarán dichas características junto con la edad de los pacientes y su género, de forma que obtendremos una matriz de 102 variables, recordando que la capa *fully connected* estaba formada de 100 neuronas. Dicha matriz constituirá los datos de entrada que aportaremos al perceptrón multicapa encargado de llevar a cabo la tarea de clasificación.

De este modo, el perceptrón estará formado por tres capas, la de entrada de 102 neuronas, una única capa oculta de 50, y finalmente, las dos neuronas de salida, que se corresponden con las clases sano y patológico, respectivamente. La activación de las neuronas de las dos primeras capas será *ReLU*, mientras que la de la última capa, *softmax*.

Finalmente, se procede a entrenar esta nueva arquitectura, con el mismo optimizador y función de pérdidas que en la Sección 6.2.1.

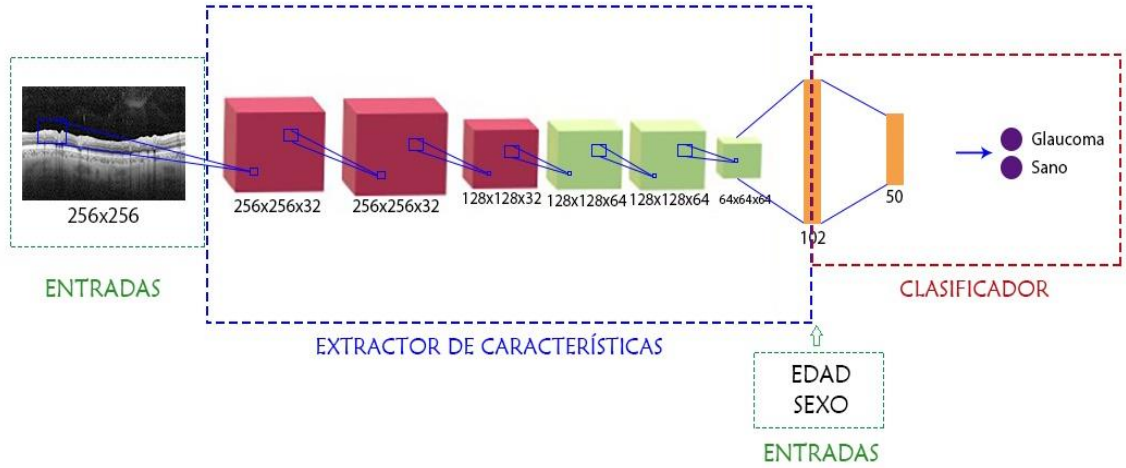


Figura 22: Arquitectura que combina un extractor de características de imágenes junto con un perceptrón multicapa.

6.2.3 Clasificador a partir de *transfer learning* y *fine tuning*

Para intentar solventar el problema del sobreajuste producido en la red diseñada desde cero, se decide hacer uso de la transferencia de conocimiento, empleando redes que ya han sido pre-entrenadas en amplias bases de datos y que disponen de conocimiento genérico. Esta tarea ha sido realizada para varias de las redes propuestas en los diferentes concursos de ImageNet, que mejores resultados obtuvieron en la sección de clasificación. Concretamente, se procede a re-entrenar las redes VGG16, VGG19, InceptionV3, Xception y ResNet aplicándoles *fine tuning*.

Todas estas arquitecturas, fueron diseñadas para trabajar con imágenes en escala de color RGB, por lo que espera recibir entradas de tres canales (una por cada color). En nuestro caso, las imágenes circumpapilares están en escala de grises, por lo que replicaremos la imagen tres veces, una para cada uno de los tres canales RGB que espera recibir.

A continuación, procederemos a mencionar las capas que han sido tuneadas de cada uno de los modelos importados:

- VGG16: en esta arquitectura, el ajuste fino se ha realizado únicamente para las últimas cuatro capas, que componen el quinto bloque. Es decir, los cuatro primeros bloques no se re-entrenan, sino que adquieren los pesos originales. El último bloque se reentrenará con las nuevas imágenes. Al reentrenar tan solo cuatro capas, estamos realizando un tuneado poco profundo.
- VGG19: esta arquitectura es similar a la anterior, solo que dispone de tres capas más de profundidad. En este caso, el último bloque está compuesto de cinco capas, que serán reentrenadas.
- InceptionV3: la técnica de *fine tuning* aplicada a esta arquitectura consiste en descongelar el último bloque, es decir, las últimas 32 capas de dicho modelo serán reentrenadas con las imágenes actuales.
- Xception: similar al caso anterior, el último bloque compuesto de 10 capas será reentrenado.
- ResNet50: en esta arquitectura, se han reentrenado los cinco últimos bloques. Dado de cada uno de ellos se compone de tres capas convolucionales, en total se han reentrenado 15 capas convolucionales.

Para finalizar, hemos de indicar las capas *fully connected* que conformarán nuestro clasificador final. Para ello, se han introducido cuatro capas: la primera, la segunda y la tercera de ellas, de 1000, 500 y 250 neuronas, respectivamente, todas ellas activadas con la función *ReLU*; mientras que la última se compone únicamente de dos neuronas, que dará como resultado la clasificación final tras aplicar la función umbral *softmax*.

Llegados a este punto, se procede a compilar y entrenar la red, para lo cual aplicaremos el optimizador de pesos SGD con una tasa de aprendizaje de $\eta = 10e - 4$. Del mismo modo que en la red diseñada desde cero, el entrenamiento consistirá en 1000 épocas, con un tamaño de *batch* de 32 muestras, para cada una de las cinco divisiones de la validación cruzada.

Capítulo 7. Resultados

Una vez entrenados y validados todos los modelos, se procede a realizar una evaluación de los resultados obtenidos, lo que nos ayudará a analizar las ventajas e inconvenientes que presenta cada uno de ellos y determinar cuál ofrece una mejor solución a nuestro problema.

7.1 Descripción de las métricas empleadas

Diversas métricas van a ser empleadas para la evaluación de los diferentes modelos. La primera de ellas se obtiene durante la fase de entrenamiento y validación de cada una de las arquitecturas. Para cada época de entrenamiento, se proporciona dos métricas, una de ellas son las pérdidas, que variarán en función del optimizador empleado y su tasa de aprendizaje, y la segunda, es la precisión de clasificación. Estas medidas se presentarán en forma de gráfico. Así, dos aspectos se han de tener en cuenta. El primero de ellos es que las pérdidas han disminuir época a época hasta acercarse a valores lo más cercanos a cero posible, mientras que la precisión ha de hacerlo a la inversa, tendiendo a la unidad. El siguiente aspecto a considerar es que las curvas de validación han de seguir el mismo patrón descrito por las de entrenamiento, lo que indicaría, por una parte, que el modelo aprende y, por otra, que no hay sobreajuste ya que los datos con los que se obtienen ambas curvas son diferentes.

Las siguientes métricas que se emplearán, serán directamente aplicadas sobre el conjunto de datos de test, es decir, sobre la predicción final de los modelos. Para ello, resulta conveniente conocer los cuatro tipos de solución que nos puede dar la clasificación:

- Verdaderos positivos: la imagen es clasificada como glaucomatosa y en la realidad presenta dicha patología.
- Verdaderos negativos: el paciente es clasificado como sano, y realmente lo está.
- Falsos positivos: se predice que el paciente padece de glaucoma, pero en la realidad no lo tiene.
- Falsos negativos: son casos en los que no se detecta ninguna patología, pero la realidad es que si la presenta.

Atendiendo a estos cuatro casos, procedemos a describir las métricas [56]:

- Precisión: hace referencia al cociente de los verdaderos positivos entre los verdaderos positivos y los falsos positivos, es decir a los pacientes en los que se detecta glaucoma y que realmente lo tienen entre el total de casos predichos como glaucoma (ya sean casos ciertos en la realidad o no).

$$\boldsymbol{Precisión} = \frac{\boldsymbol{Verdaderos\ positivos}}{\boldsymbol{Verdaderos\ positivos + Falsos\ positivos}} \quad \textbf{(7.1)}$$

- Sensibilidad, también conocida como exhaustividad o tasa de verdaderos positivos: proporciona un indicador del número de pacientes que presentan glaucoma, que han sido bien identificados.

$$\boldsymbol{Sensibilidad} = \frac{\boldsymbol{Verdaderos\ positivos}}{\boldsymbol{Verdaderos\ positivos + Falsos\ negativos}} \quad \textbf{(7.2)}$$

- F1-score: Tiene en cuenta para su cálculo, tanto la precisión como la sensibilidad:

$$\boldsymbol{F1_{score}} = \mathbf{2}\frac{\boldsymbol{Precisión * Sensibilidad}}{\boldsymbol{Precisión + Sensibilidad}} \quad \textbf{(7.3)}$$

- Curva ROC: Representa gráficamente la relación entre la tasa de verdaderos positivos y la tasa de falsos positivos, siendo esta última el cociente entre los falsos positivos y todos los pacientes sanos en la realidad.

$$\boldsymbol{Tasa\ Falsos\ Positivos} = \frac{\boldsymbol{Falsos\ positivos}}{\boldsymbol{Falsos\ positivos + Verdaderos\ negativos}} = \mathbf{1} - \boldsymbol{Especificidad} \quad (7.4)$$

- Área bajo la curva, AUC, indica el área que se recoge bajo la curva ROC. Atendiendo a la definición de dicha curva, el caso ideal sería que esta métrica resultase lo más cercana a la unidad.

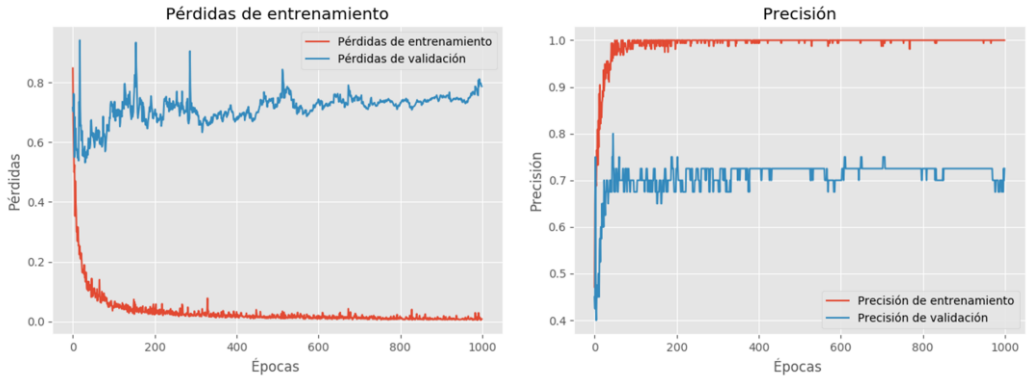
Una vez definido el procedimiento de evaluación establecido, procedemos a exponer los resultados obtenidos para los diferentes modelos.

7.2 Resultados del modelo *from scratch*.

A continuación, se muestran las gráficas que representan las pérdidas y la precisión obtenidas en la fase de entrenamiento del modelo diseñado con dos bloques convolucionales. Así, en la Gráfica 3 se observan las curvas generadas a partir del modelo inicial, es decir, sin aumentado de datos. Únicamente se ha representado los resultados obtenidos para la división número 4 de validación cruzada. Sin embargo, todos los resultados que se comenten en esta sección describen el mismo patrón para las cinco divisiones realizadas.

Prestando atención a las pérdidas, estas presentan un valor medio de en torno a 0.7 en validación, muy elevado, e incluso se puede comprobar que experimentan una evolución creciente conforme avanza el número de épocas entrenadas. Esto puede indicar que existe sobreajuste ya que las pérdidas de entrenamiento evolucionan hasta un valor prácticamente de cero, mientras que las del conjunto de validación evolucionan en sentido contrario.

Por otra parte, analizando la precisión, el modelo apenas aprende. Es cierto que el nivel de precisión es del 100% en entrenamiento, pero se mantiene constante prácticamente desde el momento inicial. En el caso de validación, la precisión que se presenta es del 72,5% y se mantiene también uniforme en el tiempo.



Gráfica 3: Resultados de la fase de entrenamiento del modelo From scratch sin aumentado de datos, para la carpeta número 4. A la izquierda se muestran las pérdidas de entrenamiento y validación. A la derecha, la precisión.

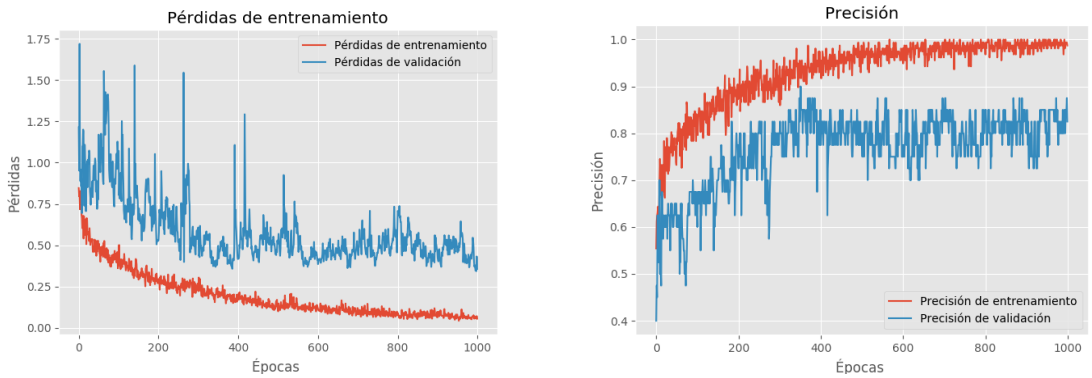
En la Tabla 8, se muestran los resultados obtenidos tras la predicción aplicada al conjunto de test, imágenes hasta el momento desconocidas por la red. Dado que se ha realizado validación cruzada, los datos de la tabla se corresponden a la media ponderada de las cinco predicciones junto con la desviación típica. Así, la precisión del modelo media es del 78.2%, mientras que la sensibilidad y F1-score son del 73.6% y 74.2%, respectivamente.

		Precisión	Sensibilidad	F1-score
Red *from-scratch*	Sin AD	0.782±0.05	0.736±0.067	0.742±0.063
	Con AD	0.814±0.06	0.758±0.115	0.764±0.111
	Con datos demográficos	0.788±0.026	0.79±0.02	0.784±0.026

Tabla 8: Resultados de precisión, sensibilidad y F1-score obtenidos en la fase de predicción del modelo *from scratch*.

Tras la obtención de estos resultados, se decide realizar un aumento sintético de datos, para disponer de un mayor número de muestras de entrenamiento e intentar, de este modo, que se produzca mayor aprendizaje. Así, en la Gráfica 4, se observan las curvas de pérdidas y validación en la fase de entrenamiento tras la creación de muestras sintéticas. En este caso, las pérdidas de la base de datos de validación describen el mismo patrón que las de entrenamiento, con una media cercana a 0.6 (menor que en el caso anterior, 0.725), si bien esta gráfica es mucho más ruidosa que en el caso previo. Las pérdidas de validación presentan muchos picos que distorsionan la señal.

En cuanto a la precisión, en este caso sí que se aprecia que la red aprende conforme evoluciona el número de épocas entrenadas. No obstante, a partir de la época 400 el aprendizaje en validación se estanca, aunque esta vez con un valor superior al caso previo, en torno al 80% de precisión (ver Tabla 8).

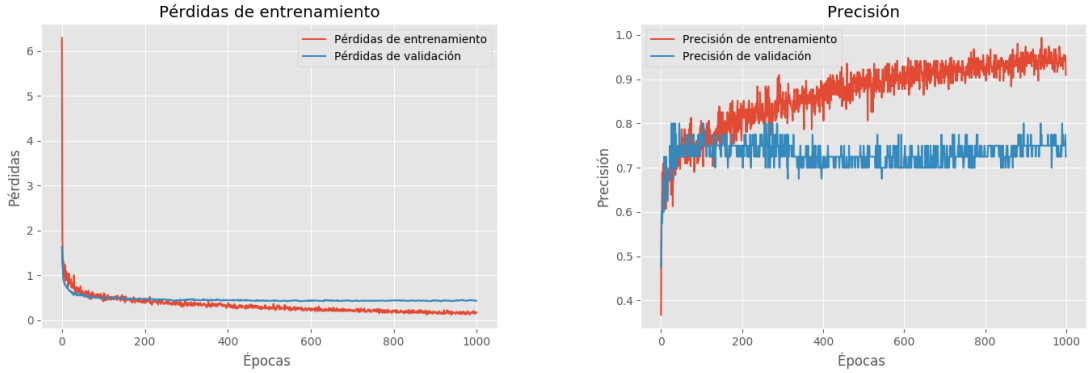


Gráfica 4: Resultados de la fase de entrenamiento del modelo From scratch con aumentado de datos, para la carpeta número 4. A la izquierda se muestran las pérdidas de entrenamiento y validación. A la derecha, la precisión.

7.3 Resultados del modelo combinado de redes neuronales convolucionales con perceptron multicapa.

Dado que el modelo *from scratch*, continúa presentando sobreajuste, y la precisión de clasificación obtenida no supera el 82% (Tabla 8), se ha decidido introducir la edad y el sexo de los pacientes al modelo de clasificación con la intención de mejorar los resultados obtenidos hasta el momento.

Analizando las curvas de pérdidas y precisión de la fase de entrenamiento para el *fold* número 4, (Gráfica 5), observamos cómo en este caso, las pérdidas son inferiores que en los casos anteriores, en concreto de en torno a 0.5, mientras que la precisión de validación media es de aproximadamente el 75%. Sin embargo, continúa presenciándose signos de sobreajuste.

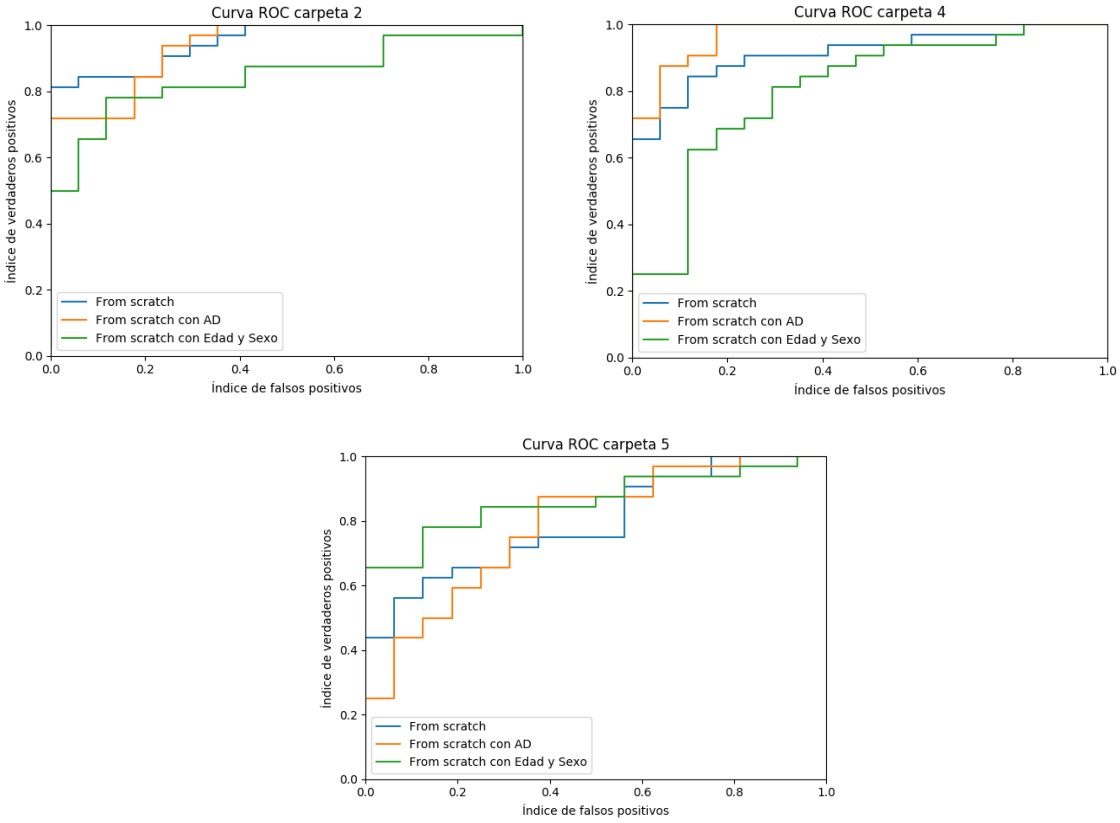


Gráfica 5: Resultados de la fase de entrenamiento del modelo *From scratch* combinado con el perceptrón multicapa, para la carpeta número 4. A la izquierda se muestran las pérdidas de entrenamiento y validación. A la derecha, la precisión.

Adicionalmente, comparando los resultados de precisión, sensibilidad y F1-score en la Tabla 8, comprobamos que estos dos últimos indicadores son superiores tras haber aportado datos adicionales al modelo e introducir el perceptrón, si bien la precisión obtenida es inferior que en el caso de aplicar aumentado sintético de datos. No obstante, la diferencia existente entre los tres modelos es muy pequeña, siendo además variables con respecto en función del *k-fold* analizado. Con respecto a esto último, cabe añadir que en el caso de introducir variables categóricas la variabilidad en los resultados obtenidos en el conjunto de los cinco *folds* es menor, lo que significa que el sobreajuste del modelo es inferior ya que para diferentes estructuras de datos, los resultados son similares y, por ello mismo, el modelo presenta mayor capacidad de generalización.

De igual modo, en la Gráfica 6 observamos que para cada *fold* la curva ROC que recoge un área más elevada bajo ella procede de diferentes modelos propuestos. Así, para el *fold* número dos, la mejor curva es la proporcionada por el modelo *from scratch* original, mientras que para los *folds* 4 y 5, las mejores curvas vienen dadas por los modelos que incorporan *data augmentation* y datos demográficos, respectivamente.

Todo ello limita la elección del modelo idóneo. El motivo puede ser debido a que, disponemos de una base de datos reducida, a pesar de que tratemos de incrementarla con datos sintéticos o variables adicionales. Además, en todos los casos se puede presenciar pequeños signos de sobreajuste.



Gráfica 6: Roc de las diferentes propuestas from scratch para los *k-folds* 2,4 y 5.

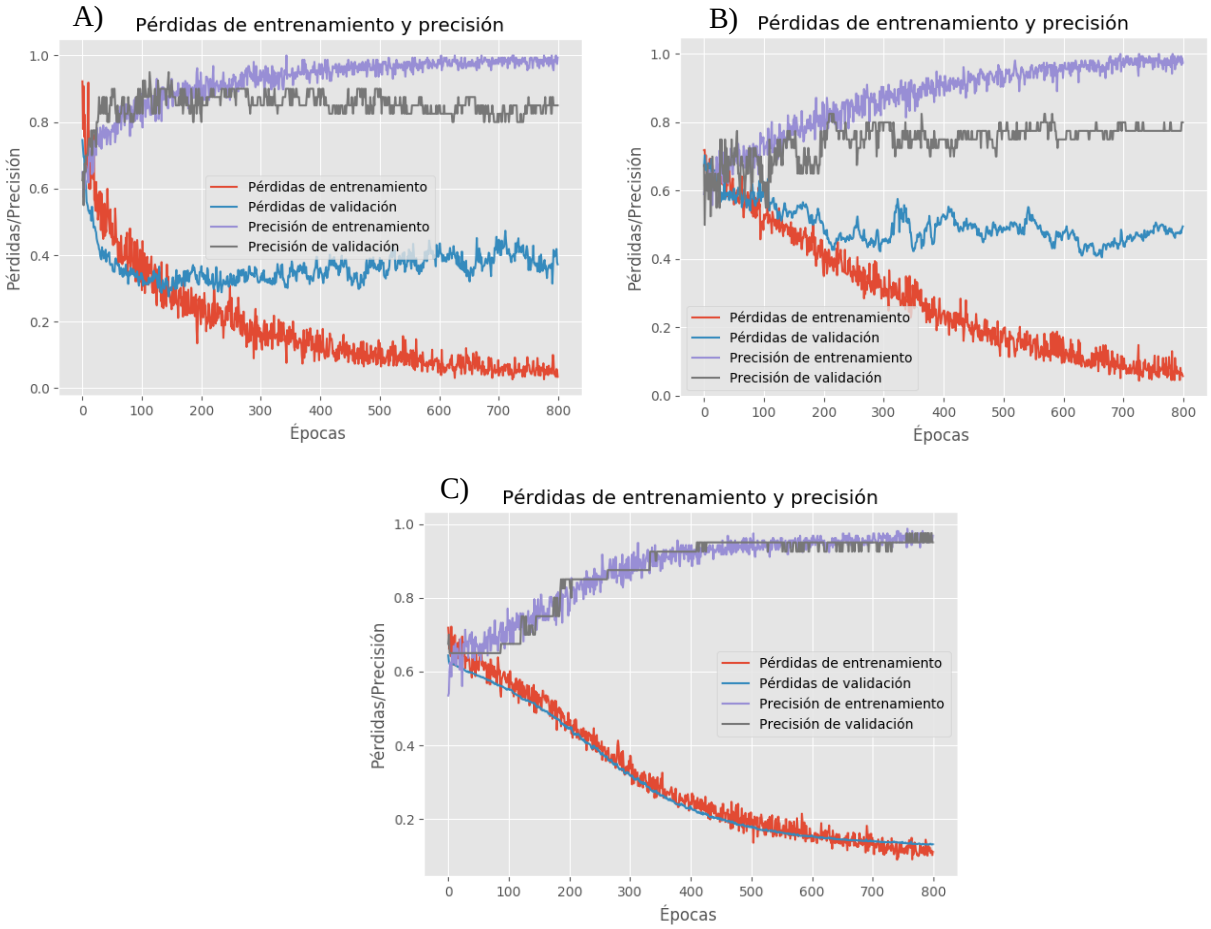
7.4 Comparativa entre modelos de *transfer learning*.

A continuación, se procede a analizar los diferentes modelos a los que se ha aplicado transferencia de conocimiento y ajuste fino. De este modo, el modelo que mayor precisión, sensibilidad, y F1-score media proporciona atendiendo a la media ponderada de los 5 *folds* de validación cruzada (ver Tabla 9), es el obtenido bajo la arquitectura VGG16, con un valor del 92.4%. Por el contrario, los peores resultados son los que se registran a la salida de la arquitectura InceptionV3. No obstante, comparando con los resultados obtenidos por la red diseñada desde cero, que recordemos, no dispone de conocimiento inicial, los cinco modelos implementados a partir de transferencia de conocimiento presentan valores en estas métricas mucho más elevados. Concretamente, la precisión de la arquitectura VGG16 experimenta un incremento del 13% con respecto al entrenamiento *from scratch* y, en el caso de la arquitectura InceptionV3, este incremento es del 4.91%.

	Precisión	Sensibilidad	F1-score
VGG16	0.924±0.034	0.924±0.034	0.924±0.034
VGG19	0.916±0.0265	0.916±0.027	0.916±0.027
InceptionV3	0.854±0.033	0.854±0.031	0.852±0.033
Xception	0.898±0.02	0.886±0.0294	0.888±0.027
ResNet50	0.88±0.011	0.87±0.018	0.868±0.019

Tabla 9: Precisión, Sensibilidad y F1-score media de los cinco *folds* para cada uno de los modelos de *transfer learning* propuestos, así como del mejor modelo *from scratch*.

Por otro lado, analizando las gráficas de pérdidas y de precisión en la fase de entrenamiento, se puede comprobar que tanto las redes InceptionV3 como Xception aparentan presentar sobreajuste tal y como se muestra en la Gráfica 7, mientras que la arquitectura VGG16 es la que aparentemente mejor aprendizaje describe.



Gráfica 7: Pérdidas y precisión de entrenamiento para *k-fold* 4. A) Xception, B) InceptionV3, C) VGG16.

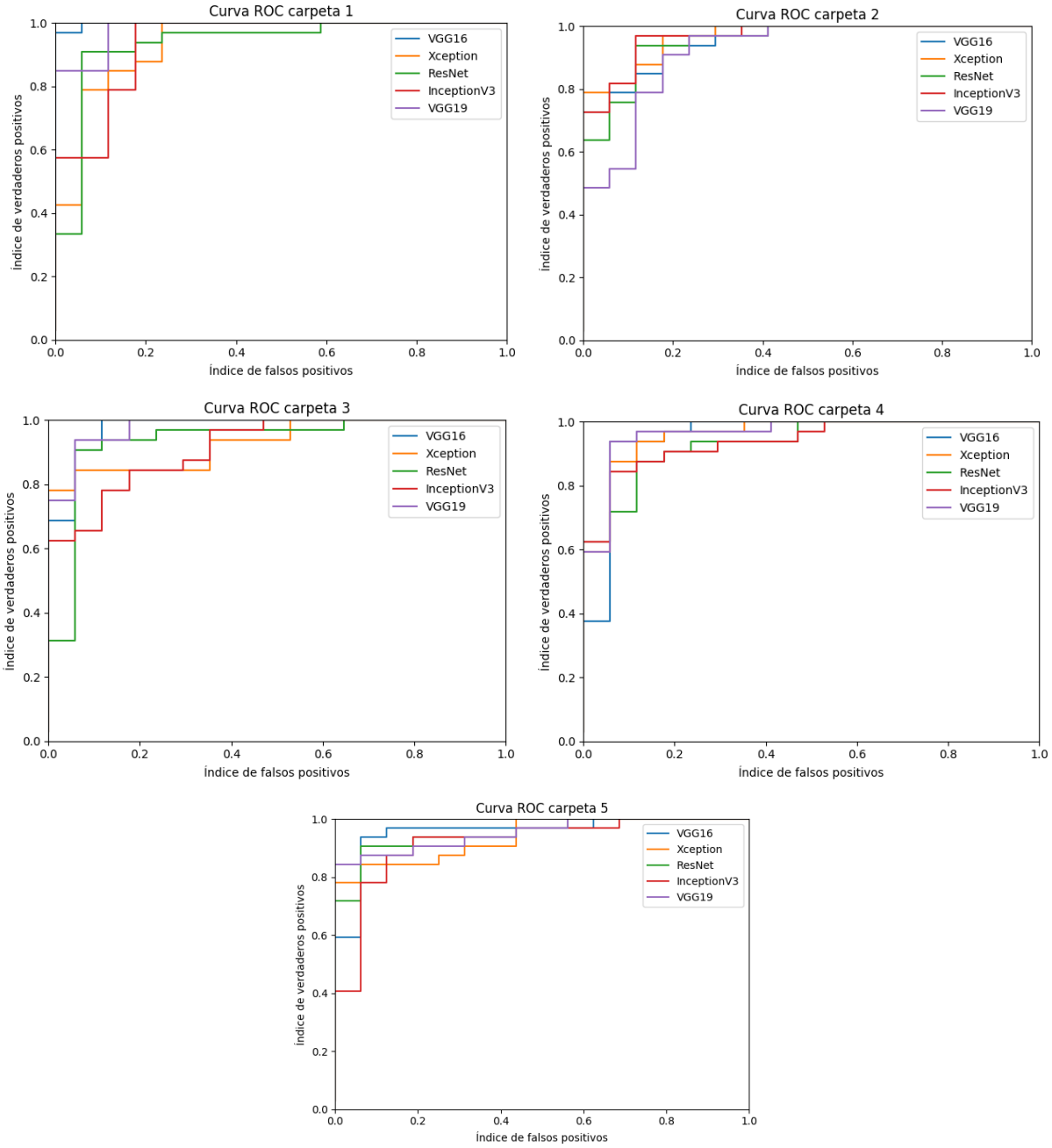
En cuanto al análisis del área recogida bajo la curva se refiere, los cinco modelos presentan un área muy elevada, en más de la mitad de los casos superior al 95% (ver Tabla 10). El hecho de que el ratio de verdaderos positivos entre falsos positivos se acerque mucho a la unidad indicaría que las patologías se están bien identificadas. Todo ello resulta favorable, ya que la idea principal perseguida en este documento es precisamente la correcta identificación y clasificación de ojos glaucomatosos.

Atendiendo a esta área en función de los diferentes *folds*, en cada uno de ellos se adopta un mayor valor en una arquitectura diferente. Así pues, para las divisiones uno y cinco, es el modelo VGG16 el que mayor área proporciona, mientras que para las divisiones 3 y 4, lo es la VGG19. También podemos indicar, que la arquitectura ResNet es la que peores resultados presenta en este tipo de análisis.

	AUC para cada *k-fold*				
	1	2	3	4	5
VGG16	0.9982	0.9554367	0.9761	0.954	0.955
Xception	0.9376	0.96513	0.93014	0.9596	0.9375
ResNet	0.9358	0.95187	0.9338	0.9338	0.91172
InceptionV3	0.9376	0.96612	0.91544	0.9375	0.91797
VGG19	0.9821746	0.9198	0.97794	0.9632	0.95117

Tabla 10: Área bajo la curva de cada uno de los modelos tras aplicarles *fine tuning*, para cada uno de los cinco *k-folds*.

Más detalladamente, analizaremos las cinco curvas ROC obtenidas, donde se puede visualizar la tasa de verdaderos positivos frente a la de falsos positivos (ver Gráfica 8). En ella, se puede observar, tal y como se apreciaba en la Tabla 10, que la arquitectura ResNet, es, de las cinco implementadas la que peores resultados proporciona. En el extremo contrario, las redes VGG16 y VGG19 son las que mejor apariencia de la curva ROC presentan, exceptuando la división 2 de datos de validación cruzada.



Gráfica 8: Curva ROC de los modelos de *transfer learning* para cada uno de los cinco *folds.*

Capítulo 8. Conclusiones y propuesta de trabajo futuro

8.1 Conclusiones y discusión.

Tras el entrenamiento y posterior predicción de los diferentes modelos planteados, se pueden extraer determinadas conclusiones. La primera de ellas y principal limitación de este problema es que disponer de bases de datos reducidas, impiden el correcto aprendizaje de las redes neuronales convolucionales. En este contexto, la aplicación de técnicas como el aumento de datos favorece a la obtención de conocimiento del mismo modo que ocurre con la incorporación de datos específicos de cada paciente que pueden contribuir a la localización y detección de glaucoma.

No obstante, para la elaboración de este documento, únicamente se dispone de datos que indican la edad y sexo de cada paciente. En este caso, a pesar de obtener mejores resultados que antes de introducirlos como parámetros de entrada además de las imágenes, la variación en la precisión del modelo es muy reducida. Esto puede ser debido a que los datos disponibles no son discriminativos, es decir, tanto los pacientes que presentan glaucoma como los que no, describen un rango de edad de más de cincuenta años, por lo que puede que resulte difícil encontrar un patrón entre una base de datos tan reducida que determine que la edad y el sexo son significativas para la detección de glaucoma.

De esta forma, el cuello de botella en el presente proyecto es la carencia de una base de datos amplia. Disponer de datos adicionales como pueden ser el campo visual de los pacientes, la presión intraocular, o datos numéricos del espesor de las capas del nervio óptico podría incrementar la precisión de nuestros modelos, ya que aportan información adicional relevante que los oftalmólogos tienen en cuenta en la práctica clínica para la determinación de glaucoma.

La transferencia de conocimiento es una gran alternativa a esta limitación ya que, como hemos comprobado, los modelos diseñados tras emplear técnicas de ajuste a arquitecturas ya entrenadas con amplias bases de datos proporcionan mejores predicciones, pues disponen de conocimiento a unos niveles que, con la carencia de datos que tenemos no podemos obtener con una red creada desde cero.

Además, analizando individualmente las arquitecturas surgidas a raíz del ILSVRC, podemos comprobar también que la red VGG16 se ajusta mejor a las características particulares de nuestro problema, además de ser ampliamente generalizable. Se ha comprobado que el empleo de arquitecturas profundas mejora el aprendizaje, estableciendo una analogía entre los resultados obtenidos con nuestra red diseñada desde cero de apenas ocho capas de profundidad (Tabla 8), y la mejor arquitectura de *transfer learning*, 16 capas de profundidad (Tabla 9). Sin embargo, el incremento de la profundidad es favorecedor únicamente hasta un determinado número de capas, ya que redes más profundas como la VGG19, Xception e InceptionV3, proporcionan peores resultados de clasificación. No obstante, el coste computacional requerido por la arquitectura VGG16 es muy elevado, lo que requiere de disponer de grandes memorias así como un mayor tiempo invertido en la fase de entrenamiento.

Con todo ello, ha sido posible obtener un sistema de clasificación basado en imágenes que nos permite discernir, con alto grado de precisión entre pacientes sanos y patológicos. Esto supone incrementar la velocidad de los sistemas actuales de detección de glaucoma, ya que como se vio en la sección 5.7, los modelos existentes hasta el momento que recurren a la inteligencia artificial para la detección de dicha patología, requieren de un trabajo previo de preparación de datos, tales como segmentación de capas o creación de mapas de probabilidades.

8.2 Trabajo futuro

Tras la realización de este proyecto, y el análisis de resultados y conclusiones obtenidas, se puede establecer una guía de actuaciones futuras que permitan contribuir de algún modo a la mejora de los resultados de clasificación alcanzados, dado que se trata de un problema de diagnóstico que requiere de gran precisión.

Dada la carencia de datos disponible, sería conveniente conformar una base de datos sólida, con mayor número de imágenes y un consistente balanceo entre clases. De esta forma el hecho de dotar de más datos de partida a la red le permitiría alcanzar un mayor nivel de conocimiento, por lo que podría favorecer a la mejora en los resultados de clasificación y, dado que se incrementaría el número de datos de entrenamiento podría adquirir una mayor capacidad de generalización evitando así la aparición de sobreajuste.

Además, en el caso de disponer de clases balanceadas se proporcionaría mayor información acerca de los pacientes con glaucoma dado que esta es la clase más afectada por el desbalanceo. Este hecho permitiría alcanzar un mayor nivel de abstracción de características en lo que a la detección de glaucoma a partir de imágenes circumpapilares se refiere, que favorecerían a la correcta identificación de dicha patología.

Por otra parte, dada la mejora que en el modelo *from scratch* supone la incorporación de datos demográficos, sería conveniente introducir a los modelos nuevas variables categóricas e historiales de paciente tales como la presión intraocular, el campo visual, resultados de paquimetría, antecedentes familiares y trastornos oculares como la miopía e hipermetropía, entre otros, ya que en la práctica clínica son datos estudiados por el especialista en el momento de discernir entre ojos sanos y glaucomatosos. Este hecho puede aportar mayor información en lo referente a la identificación de glaucoma que permitiría disminuir el número de falsos positivos y falsos negativos obtenidos hasta el momento.

En el caso de que todo lo anterior fuera posible y dispusiéramos de un amplio número de muestras, se podría dotar a la red implementada desde cero de mayor profundidad para comprobar si, tal y como habíamos previsto con la implementación de la estructura VGG16, la profundidad de la arquitectura contribuye al aprendizaje y mejora de resultados de clasificación proporcionados.

Referencias

[1] "Glaucoma – Advanced, LAbel-free High resolution Automated OCT Diagnostics." [Online]. Disponible: https://galahad-project.eu/.

[2] C. Urtubia Vicario, *Neurobiología de la visión*. Edicions UPC, 1999.

[3] "Estructura del Ojo Humano - Óptica Meseguer Xàtiva." [Online]. Disponible: http://opticameseguer.com/estructura-del-ojo-humano/.

[4] G. K. Lang, *Oftalmología : texto y atlas en color*. Masson, 2006.

[5] J. J. Kanski and B. (Bradley) Bowling, *Kanski. Oftalmología clínica : Un enfoque sistemático*, Séptima edición. 1976.

[6] "OMS | Ceguera," *WHO*, 2017. [Online]

[7] J. C. Javitt, A. M. McBean, G. A. Nicholson, J. D. Babish, J. L. Warren, and H. Krakauer, "Undertreatment of Glaucoma among Black Americans," *N. Engl. J. Med.*, vol. 325, no. 20, pp. 1418–1422, Nov. 1991.

[8] "Glaucoma: ¿qué es? | Clínica Villoria, oftalmología avanzada." [Online]. Disponible: https://clinicavilloria.es/otros-tratamientos/glaucoma/.

[9] J. M. Schmitt, "Optical coherence tomography (OCT): a review," *IEEE J. Sel. Top. Quantum Electron.*, vol. 5, no. 4, pp. 1205–1215, 1999.

[10] F. Lavinsky, G. Wollstein, J. Tauber, and J. S. Schuman, "The Future of Imaging in Detecting Glaucoma Progression," *Ophthalmology*, vol. 124, no. 12, pp. S76–S82, Dic. 2017.

[11] C. Griño García-Pardo, F. Lugo Quintás, M. León, S. Ligero, J. Ruiz-Moreno, and J. Montero Moreno, "Tomografía de Coherencia Óptica (OCT) Funcionamiento y utilidad en patología macular (I)", 2008.

[12] J. Moor, "The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years," *AI Mag.*, vol. 27, no. 4, pp. 87–87, Dic. 2006.

[13] X. Wang, M. Yang, S. Zhu, and Y. Lin, "Regionlets for Generic Object Detection," 2013.

[14] B. Ko, "A Brief Review of Facial Emotion Recognition Based on Visual Information," *Sensors*, vol. 18, no. 2, p. 401, Ene. 2018.

[15] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to Sequence Learning with Neural Networks." pp. 3104–3112, 2014.

[16] F. Jiang *et al.*, "Artificial intelligence in healthcare: past, present and future," *Stroke Vasc. Neurol.*, vol. 2, no. 4, pp. 230–243, Dic. 2017.

[17] M. . Gardner and S. . Dorling, "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences," *Atmos. Environ.*, vol. 32, no. 14–15, pp. 2627–2636, Ago. 1998.

[18] A. N. Ramesh, C. Kambhampati, J. R. T. Monson, and P. J. Drew, "Artificial intelligence in medicine.," *Ann. R. Coll. Surg. Engl.*, vol. 86, no. 5, pp. 334–8, Sep. 2004.

[19] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dic. 2017.

[20] A. Kumar, J. Kim, D. Lyndon, M. Fulham, and D. Feng, "An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification," *IEEE J. Biomed. Heal. Informatics*, vol. 21, no. 1, pp. 31–40, Ene. 2017.

[21] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*, 2014, pp. 844–848.

[22] D. Svozil, V. Kvasnicka, and J. Pospichal, "Introduction to multi-layer feed-forward neural networks," *Chemom. Intell. Lab. Syst.*, vol. 39, no. 1, pp. 43–62, Nov. 1997.

[23] S. K. Pal and S. Mitra, "Multilayer perceptron, fuzzy sets, and classification," *IEEE Trans. Neural Networks*, vol. 3, no. 5, pp. 683–697, 1992.

[24] Y. Lecun and Y. Bengio, "Convolutional Networks for Images, Speech, and Time-Series," 1995.

[25] "A Comprehensive Guide to Convolutional Neural Networks." [Online]. Disponible: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.

[26] V. Dumoulin, F. Visin, and G. E. P. Box, "A guide to convolution arithmetic for deep learning," 2018.

[27] B. Karlik and A. V. Olgac, "Performance Analysis of Various Activation Functions in Generalized MLP Architectures of Neural Networks," *Int. J. Artif. Intell. Expert Syst.*, vol. 1, no. 4, pp. 111–122, 2011.

[28] "CS231n Convolutional Neural Networks for Visual Recognition." [Online]. Disponible: http://cs231n.github.io/neural-networks-1/.

[29] Y. Cui, F. Zhou, J. Wang, X. Liu, Y. Lin, and S. Belongie, "Kernel Pooling for Convolutional Neural Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3049–3058.

[30] S. Ruder, "An overview of gradient descent optimization algorithms *," Dublin, 2017.

[31] L. Taylor and G. Nitschke, "Improving Deep Learning using Generic Data Augmentation," 2017.

[32] N. Srivastava, G. Hinton, A. Krizhevsky, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," 2014.

[33] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *undefined*, 2015.

[34] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How Does Batch Normalization Help Optimization?," May 2018.

[35] K. A. Ross *et al.*, "Cross-Validation," in *Encyclopedia of Database Systems*, Boston, MA: Springer US, 2009, pp. 532–538.

[36] N. Tajbakhsh *et al.*, "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1299–1312, May 2016.

[37] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," 2015.

[38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks." pp. 1097–1105, 2012.

[39] C. Szegedy *et al.*, "Going deeper with convolutions," 2014.

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2015.

[41] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2014.

[42] C. Szegedy, V. Vanhoucke, S. Ioffe, and J. Shlens, "Rethinking the Inception Architecture for Computer Vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.

[43] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1800–1807.

[44] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[45] E. J. Carmona Suárez, "Tutorial sobre Máquinas de Vectores Soporte (SVM)," Madrid, 2014.

[46] R. Agrawal, "K-Nearest Neighbor for Uncertain Data," 2014.

[47] S. J. Kim, K. J. Cho, and S. Oh, "Development of machine learning models for diagnosis of glaucoma," *PLoS One*, vol. 12, no. 5, p. e0177726, May 2017.

[48] F. R. Silva, V. G. Vidotti, F. Cremasco, M. Dias, E. S. Gomi, and V. P. Costa, "Sensitivity and specificity of machine learning classifiers for glaucoma diagnosis using Spectral Domain OCT and standard automated perimetry," *Arq. Bras. Oftalmol.*, vol. 76, no. 3, pp. 170–174, Jun. 2013.

[49] K. A. Barella, V. P. Costa, V. Gonçalves Vidotti, F. R. Silva, M. Dias, and E. S. Gomi, "Glaucoma Diagnostic Accuracy of Machine Learning Classifiers Using Retinal Nerve Fiber Layer and Optic Nerve Data from SD-OCT.," *J. Ophthalmol.*, vol. 2013, p. 789129, 2013.

[50] Q. Abbas, "Glaucoma-Deep: Detection of Glaucoma Eye Disease on Retinal Fundus Images using Deep Learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 6, 2017.

[51] X. Chen, Y. Xu, D. W. Kee Wong, T. Y. Wong, and J. Liu, "Glaucoma detection based on deep convolutional neural network," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 715–718.

[52] S. K. Devalla *et al.*, "A Deep Learning Approach to Digitally Stain Optical Coherence Tomography Images of the Optic Nerve Head," *Investig. Opthalmology Vis. Sci.*, vol. 59, no. 1, p. 63, Jan. 2018.

[53] H. Muhammad *et al.*, "Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects," *J. Glaucoma*, vol. 26, no. 12, pp. 1086–1094, 2017.

[54] "Keras Documentation." [Online]. Disponible: https://keras.io/.

[55] "scikit-learn: machine learning in Python — scikit-learn 0.21.2 documentation." [Online]. Disponible: https://scikit-learn.org/stable/.

[56] Y. Ding *et al.*, "A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using ^{18} F-FDG PET of the Brain," *Radiology*, vol. 290, no. 2, pp. 456–464, Feb. 2019.

Anexo I

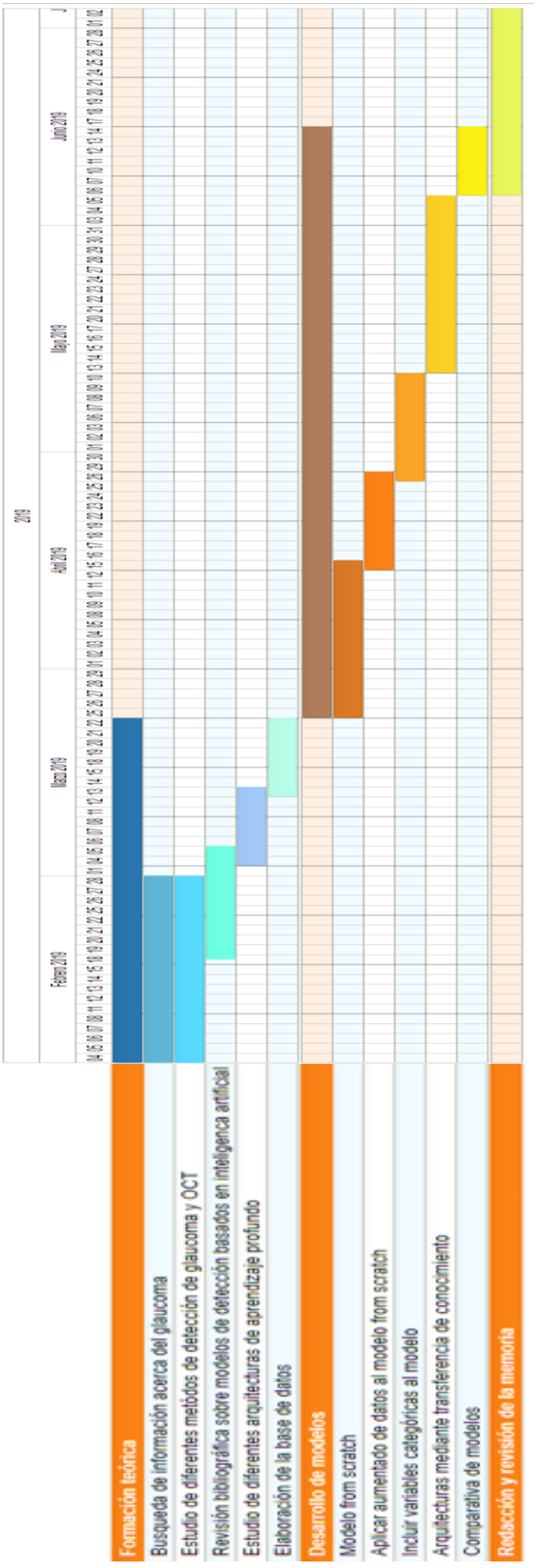


Figura 23: Diagrama de Grantt basado en el reparto de tareas llevado a cabo en la realización del trabajo fin de grado