



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

— **TELECOM** ESCUELA
TÉCNICA **VLC** SUPERIOR
DE **UPV** INGENIEROS DE
TELECOMUNICACIÓN

DISCRIMINACIÓN DE ENFERMEDADES OFTALMOLÓGICAS MEDIANTE REDES NEURONALES CONVOLUCIONALES: MAPAS DE ACTIVACIÓN PARA LA LOCALIZACIÓN DE TEJIDO PATOLÓGICO EN IMÁGENES DE FONDO DE OJO

Pablo González Carrizo

Tutor: Valery Naranjo Ornedo

Cotutor: Adrián Colomer Granero

Trabajo Fin de Máster presentado en la Escuela Técnica Superior de Ingenieros de Telecomunicación de la Universitat Politècnica de València, para la obtención del Título de Máster en Ingeniería Telecomunicación

Curso 2018-19

Valencia, 9 de septiembre 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



Dedicado a Fran, porque aún tenemos que hablar de muchas cosas, compañero del alma.

Abstract

Diabetic Retinopathy and Age-Related Macular Degeneration are two of the main causes of blindness all around the globe. However, both diseases can be treated, and their effects minimized, if they are detected on early stages. The current screening system is not scalable for the unstoppable growth of both diseases. During this work, three different automatic screening systems are proposed. These systems use some convolutional neural networks previously trained on different domains. The proposed models are more robust and more interpretable than some of the state-of-the-art ones through the use of more than 39000 images for training and the use of activation maps during predictions.

Resumen

La Retinopatía Diabética y la Degeneración Macular Asociada a la Edad son dos de las principales causas de ceguera en todo el mundo. Sin embargo estas enfermedades pueden ser tratadas, y sus efectos minimizados, si son detectadas a tiempo. El sistema actual, basado en la revisión manual de expertos no es viable ante el crecimiento imparable de ambas. En este trabajo se proponen 3 sistemas para su detección automática. Estos sistemas usan como punto de partida para su entrenamiento redes neuronales convolucionales entrenadas previamente en otros dominios. Las principales ventajas de los sistemas propuestos frente a los modelos del estado del arte analizados es la robustez que proporciona haber sido entrenados con más de 39000 imágenes y la gran interpretabilidad conseguida gracias al uso de mapas de activación en las predicciones.

Índice general

Abstract

Resumen	I
---------	---

Lista de figuras	
------------------	--

Lista de tablas	III
-----------------	-----

Abreviaciones	IV
---------------	----

1. Introducción	1
------------------------	---

1.1. Motivación	1
1.2. Objetivos	4
1.3. Principales contribuciones	4
1.4. Estructura	5

2. El ojo y sus patologías	7
-----------------------------------	---

2.1. Anatomía y fisiología ocular	7
2.1.1. La retina y su importancia	10
2.2. Principales patologías de la retina	13
2.2.1. Retinopatía Diabética	15
2.2.2. Degeneración macular asociada a la edad	18
2.2.3. Sistemas de diagnóstico	20

3. Machine Learning y aplicaciones médicas	23
---	----

3.1. IA, Big Data, Machine Learning y Deep Learning	25
3.2. Redes neuronales, descenso de gradiente y backpropagation .	28
3.2.1. Redes neuronales convolucionales	33
3.3. Transfer Learning	38

3.3.1. Transfer Learning con imágenes	38
3.4. Explicabilidad las redes convolucionales	41
3.5. Aplicaciones médicas del Machine Learning	42
3.5.1. Pronóstico	44
3.5.2. Diagnóstico	44
3.5.3. Tratamiento	45
3.5.4. Retos clave	45
3.6. Correlación no implica causalidad	46
4. Estado del arte en detección de RD y DMAE	49
4.1. Aproximaciones basadas en Machine Learning	50
4.1.1. Detección de RD mediante Machine Learning	50
4.1.2. Detección de DMAE mediante Machine Learning	53
4.2. Aproximaciones basadas en Deep Learning	55
4.2.1. Detección de RD mediante Deep Learning	56
4.2.2. Detección de DMAE mediante Deep Learning	57
5. Diseño de Sistema de Detección de RD y DMAE	59
5.1. Exploración de los datos	59
5.2. Recursos utilizados	62
5.3. Pre-procesado de las imágenes	65
5.4. Diseño del sistema 1: Gran clasificador	67
5.5. Diseño del sistema 2: Clasificador Multietapa	68
5.6. Diseño del sistema 3: Ensemble de Clasificadores	71
5.7. Diseño del Sistema de Predicción e Interpretación	73
6. Análisis de los resultados obtenidos	74
6.1. Evaluación de sistemas de Machine Learning	75
6.2. Evaluación del Sistema 1: Gran Clasificador	76
6.3. Evaluación del Sistema 2: Clasificador Multietapa	77
6.3.1. Etapa 1: Clasificador Sano/Enfermo	77
6.3.2. Etapa 2: Clasificador RD/DMAE	84
6.4. Evaluación del Sistema 3: Ensemble de Clasificadores	86
6.5. Sistema de Predicción e Interpretación	89
7. Conclusiones	96

7.1. Trabajo futuro	97
Referencias	99

Lista de figuras

1.1.	Prevalecencia y previsión de crecimiento de la diabetes y la RD a nivel mundial. Gráfico de elaboración propia	2
1.2.	Modelo de cámara de fondo de ojo Eidon de la compañía Centervue	3
2.1.	Estructura del ojo humano. Fuente: Wikipedia	9
2.2.	Elementos de la retina. Fuente: Kaggle (anotaciones de elaboración propia)	12
2.3.	Efectos en la visión de las enfermedades analizadas en este trabajo: (a) Visión normal, (b) con Retinopatía Diabética no proliferativa, (c) con Retinopatía Diabética proliferativa y (d) con degeneración macular asociada a la edad. Fuente: American Academy of Ophtalmology (www.aao.org)	14
2.4.	Lesiones típicas de la Retinopatía Diabética. Elaboración proia	17
2.5.	Ejemplo de retina en la que se ha producido neovascularización	17
2.6.	Retina con drusas a causa de DMAE	20

3.1. Interés, a lo largo del tiempo y en todo el mundo, del término Machine Learning en el buscador Google. Datos de Enero de 2014 a Julio de 2019. Un valor de 100 indica la máxima popularidad del término. Los valores 50 y 0 indican, respectivamente, que un término es la mitad de popular en relación con el valor máximo o que no existen suficientes datos del término. Fuente de los datos: Google Trends	25
3.2. El Machine Learning es un campo perteneciente a la Inteligencia Artificial. El Deep Learning, a su vez, es un campo dentro del Machine Learning. Elaboración propia	27
3.3. Interés, a lo largo del tiempo y en todo el mundo, de los término Machine Learning (en azul), Deep Learning (en rojo) y Big Data (en amarillo) en el buscador Google. Datos de Enero de 2014 a Julio de 2019. Un valor de 100 indica la máxima popularidad del término. Los valores 50 y 0 indican, respectivamente, que un término es la mitad de popular en relación con el valor máximo o que no existen suficientes datos del término. Fuente de los datos: Google Trends	28
3.4. Representación de una red neuronal con dos capas ocultas. Cada uno de los círculos representa una neurona. Elaboración propia	29
3.5. Representación de una sola neurona con 3 entradas. Cada una de esas entradas tiene asociado un peso. La neurona utiliza la función de activación ReLU. Elaboración propia	29
3.6. Resultado de la convolución de una imagen con un filtro Sobel de 3x3 horizontal (arriba) y otro vertical (abajo) Fuente: https://victorzhou.com/blog/intro-to-cnns-part-1/	35
3.7. Representación del proceso de Max Pooling con un filtro de 2x2 sobre una imagen de 4x4. Elaboración propia	35
3.8. Representación de los mapas de activación de una red convolucional con 2 capas convolucionales y 3 fully-connected. Cada capa convolucional va seguida de una de max pooling. Fuente: http://scs.ryerson.ca/~aharley/vis/conv/flat.html	36

3.9. Mapas de atención generados por Grad-Cam para distintas clases de Imagenet. Fuente: https://github.com/raghakot/keras-vis	41
3.10. Diferencias entre Software tradicional y Machine learning. Elaboración propia.	43
3.11. Correlación entre el aumento de la temperatura media global y el descenso del número de piratas. Fuente: https://www.jotdown.es/2016/06/correlacion-no-implica-causalidad/	48
3.12. Correlación entre el número de ahogados en piscinas de Estados Unidos y el número de apariciones en películas de Nicholas Cage. Fuente: http://www.tylervigen.com/spurious-correlations	48
4.1. Imagen binaria de los vasos sanguíneos de la retina. Fuente: http://www.aria-database.com/	52
4.2. Resumen de las fases de la detección de DMAE mediante Machine Learning. Fuente: (Pead et al. 2019)	54
5.1. Logos de las dos librerías de Python utilizadas para la investigación: Keras y Tensorflow.	64
5.2. Ejemplos de Jupyter Notebooks. Fuente: https://jupyter.org/	64
5.3. Arquitectura utilizada para el sistema 1. Elaboración propia	68
5.4. Arquitectura del sistema clasificador en dos etapas Elaboración propia	69
5.5. Arquitectura utilizada para los clasificadores de ambos subsistemas del segundo diseño.	71
5.6. Arquitectura del sistema de ensemble de clasificadores simples. Los bloques superiores representan los conjuntos de imágenes utilizados para el entrenamiento de los mismos Elaboración propia	72
6.1. Pérdidas para el dataset de validación del entrenamiento de los bloques 2,3,4 y 5 (y FC) del clasificador Sano/Enfermo. Se ha aplicado un filtro de suavizado.	80

6.2.	Salida de la primera etapa del Sistema Multietapa para una imagen de una retina enferma de RD	81
6.3.	Salida de la primera etapa del Sistema Multietapa para una imagen de una retina enferma de DMAE	82
6.4.	Salida de la primera etapa del Sistema Multietapa para una imagen de una retina sana	83
6.5.	Salida de la segunda etapa del Sistema Multietapa para una imagen de una retina enferma de RD	85
6.6.	Salida de la segunda etapa del Sistema Multietapa para una imagen de una retina enferma de DMAE	85
6.7.	Salida del Sistema 3 para una imagen de una retina enferma de RD (omitidos los mapas de activación)	88
6.8.	Salida del Sistema 3 para una imagen de una retina enferma de DMAE (omitidos los mapas de activación)	88
6.9.	Salida del Sistema 3 para una imagen de una retina sana (omitidos los mapas de activación)	89
6.10.	Respuesta del Sistema de Predicción a la imagen de una retina con DMAE. Etapa primera del Sistema Multietapa	90
6.11.	Respuesta del Sistema de Predicción a la imagen de una retina con DMAE. Etapa segunda del Sistema Multietapa	90
6.12.	Respuesta del Sistema de Predicción a la imagen de una retina con RD. Etapa primera del Sistema Multietapa	91
6.13.	Respuesta del Sistema de Predicción a la imagen de una retina con RD. Etapa segunda del Sistema Multietapa	91
6.14.	Respuesta del Sistema de Predicción a la imagen de una retina con DMAE. Mapa de activación del primer clasificador del Ensemble de Clasificadores	92
6.15.	Respuesta del Sistema de Predicción a la imagen de una retina con DMAE. Mapa de activación del segundo clasificador del Ensemble de Clasificadores	92
6.16.	Respuesta del Sistema de Predicción a la imagen de una retina con DMAE. Mapa de activación del tercer clasificador del Ensemble de Clasificadores	93

6.17. Respuesta del Sistema de Predicción a la imagen de una retina sana. Mapa de activación del primer clasificador del Ensemble de Clasificadores	94
6.18. Respuesta del Sistema de Predicción a la imagen de una retina sana. Mapa de activación del segundo clasificador del Ensemble de Clasificadores	94
6.19. Respuesta del Sistema de Predicción a la imagen de una retina sana. Mapa de activación del tercer clasificador del Ensemble de Clasificadores	95

Lista de tablas

2.1. Niveles de gravedad de la Retinopatía Diabética en función de las lesiones observadas	18
5.1. Cantidad de imágenes de cada tipo en cada uno de los conjuntos de imágenes utilizados	60
5.2. Cantidad de imágenes de cada tipo en el conjunto completo de datos utilizado	61
5.3. Características de las imágenes de cada uno de los conjuntos utilizados	62
5.4. Distribución de las imágenes del clasificador Sano/Enfermo del sistema 2.	69
6.1. Resultados del entrenamiento para distintos batch size. Modelos evaluados con el dataset de validación	78
6.2. Resultados del entrenamiento para distintos learning rate. Modelos evaluados con el dataset de validación	78
6.3. Resultados del entrenamiento para distintos bloques convolucionales entrenados. Modelos evaluados con el dataset de validación	79
6.4. Resultados del entrenamiento de la segunda etapa del Sistema Multietapa. Modelos evaluados con el dataset de validación	84

6.5. Resultados del entrenamiento de la segunda etapa del Sistema Multietapa. Modelos evaluados con el dataset de validación	84
6.6. Resultados del entrenamiento del sistema 3. Modelos evaluados con el dataset de validación	86
6.7. Resultados del entrenamiento con diferentes arquitecturas del sistema 3. Modelos evaluados con el dataset de validación . .	87

Abreviaciones

RD	Retinopatía Diabética
DMAE	Degeneración Macular Asociada a la Edad
NPDR	Non-Proliferative Diabetic Retinopathy
PDR	Proliferative Diabetic Retinopathy
ARIA	Automated Retinal Image Analyzer
IA	Inteligencia Artificial
CNN	Convolutional Neural Network
SVM	Support Vector Machine

Capítulo 1

Introducción

Durante este capítulo inicial se presenta el contexto y la motivación principal detrás de este trabajo, los objetivos perseguidos y la estructura en la que se plasma toda esta información a lo largo del mismo.

El presente documento pretende mostrar todas las tareas de investigación realizadas para la realización del **Trabajo Final de Máster** que permite la obtención del título de **Máster Universitario en Ingeniería de Telecomunicaciones** de la Universidad Politécnica de Valencia. Este trabajo supone 30 créditos ECTS (de los 120 créditos totales de la titulación), lo que equivale aproximadamente a 750 horas de trabajo.

1.1. Motivación

La Organización Mundial de la Salud (OMS) estima que, en 2010, 285 millones de personas padecían algún tipo de discapacidad visual. De ellas, 39 millones eran ciegas. (WHO & others 2013). El informe detallaba 7 principales causas de discapacidad visual entre las que se encontraban las 2 enfermedades que se analizarán en este trabajo: la **Retinopatía Diabética** y la **Degeneración Macular Asociada a la Edad**. Según se estimaba, el 80 % de estas discapacidades podrían haberse evitado con las intervenciones adecuadas para su prevención. En respuesta, la OMS lanzaba su plan de

acción que comenzaría en 2014 y finalizaría en 2019. El informe¹ asociado a este plan de acción ponía de manifiesto la necesidad de que los servicios de salud ocular se convirtieran en parte integral del sistema primario de salud y se resaltaba la importancia de las campañas de prevención.

La Retinopatía Diabética (RD) pertenece al grupo de las enfermedades vasculares, y se ha convertido en la **principal causa evitable de ceguera en todo el mundo**. Esta patología se da actualmente en el 35 % de las personas con diabetes, enfermedad que afecta al 8.5 % de la población mundial (IAPB 2016), (IDF 2017) . Se estima que 191 millones de personas sufrirán retinopatía diabética en 2030 (Yingfeng Zheng et al. 2012). La incidencia de la RD es del 50 % a partir de los 10 años de la aparición de la diabetes, y del 90 % a partir de los 30 años (Mookiah, U Rajendra Acharya, Chua, et al. 2013).

En la Figura 1.1 se puede observar la previsión esperada de crecimiento entre 2015 y 2040, tanto en el número de casos de diabetes, como en el de casos de diabetes que dan lugar a RD.² El aumento de la población mundial, y el envejecimiento de la misma serán factores determinantes en este crecimiento, pero también lo serán el aumento de casos de sobrepeso y la vida sedentaria.

La diabetes supone, aproximadamente, el 11.6 % del presupuesto total de salud de la mayoría de países (Zhang et al. 2009). Además, el coste de los pacientes con RD supera notablemente al de los pacientes sin dicha patología,

¹<https://www.who.int/blindness/actionplan/en/>

²Datos de <https://atlas.iapb.org/vision-trends/diabetic-retinopathy>

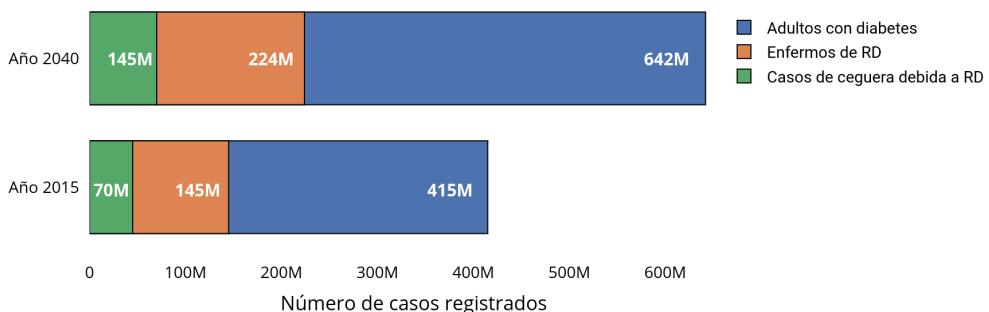


Figura 1.1: Prevalecencia y previsión de crecimiento de la diabetes y la RD a nivel mundial.
Gráfico de elaboración propia

incrementándose éste con la gravedad de la RD (Zhang et al. 2017).

Es importante destacar el hecho de que casi el 75 % de las personas que sufren Retinopatía Diabética pertenecen a países en vías de desarrollo (Mansour 2017), donde no existen los medios adecuados para su detección temprana ni su tratamiento.

Por otro lado, la **Degeneración Macular Asociada a la Edad (DMAE)** es la más común de las enfermedades que afectan a la retina. Esta patología, de tipo degenerativo, es la **mayor causa de ceguera en países desarrollados**, dándose en un 9 % de la población mundial (Wong et al. 2014). Hasta el 80 % de los casos de ceguera causados por esta enfermedad son evitables si son detectados y tratados a tiempo. (Pascolini & Mariotti 2012)

El rápido crecimiento de estas enfermedades hace insostenible el sistema actual basado únicamente en la revisión de expertos. Es necesario introducir en las clínicas sistemas de detección automática a partir de imágenes digitales que permitirían agilizar el trabajo de los médicos o incluso permitir el diagnóstico en zonas donde ni siquiera existen ese tipo de expertos. Aunque existen diferentes métodos para el diagnóstico como la tomografía de coherencia óptica (TCO) o la angiografía, el método más utilizado actualmente se basa en el análisis de **imágenes de fondo de ojo** obtenidas mediante cámaras especializadas (Figura 1.2). Este tipo de análisis se ha impuesto al resto de métodos por la facilidad de uso de las cámaras y su menor coste.

A la fecha de publicación de este trabajo (Septiembre de 2019) aún se des-



Figura 1.2: Modelo de cámara de fondo de ojo Eidon de la compañía Centervue

conoce cuál ha sido el grado de eficacia del plan de acción propuesto por la OMS, cuyo objetivo principal era la reducción de un 25 % de los casos de discapacidad visual evitables. Lo que sí que se ha podido comprobar es el crecimiento experimentado en el número de investigaciones realizadas en torno a la detección automática de algunas de estas enfermedades, entrando a la obra nuevos actores como Google que han permitido dar pasos de gigante en la lucha contra este tipo de patologías.³

1.2. Objetivos

El objetivo principal de esta investigación ha sido el **desarrollo de un sistema de detección automática de Retinopatía Diabética y Degeneración Macular Asociada a la Edad**. Sin embargo, al ser este un objetivo amplio, se han establecido una serie de objetivos más específicos que se detallan a continuación:

- Estudio de la anatomía y fisiología del ojo humano, enfocándose en las causas y los efectos de las enfermedades analizadas.
- Análisis y comparación de las principales aproximaciones a la detección automática de ambas patologías realizadas hasta la fecha, tanto las basadas en Machine Learning como en Deep Learning.
- Diseño, desarrollo y evaluación de diversas topologías de redes neuronales convolucionales en la detección de ambas patologías
- Interpretación de las redes convolucionales, tratando de comprender qué factores le han ayudado a predecir, en cada caso, la existencia o ausencia de la enfermedad.

1.3. Principales contribuciones

Las principales contribuciones de este trabajo giran en torno a dos características: la **robustez** y la **interpretabilidad**.

³<https://ai.googleblog.com/2018/12/improving-effectiveness-of-diabetic.html>

- En busca de la **robustez** se han utilizado más de 39000 imágenes procedentes de 13 datasets distintos para el entrenamiento de los modelos. La combinación de las predicciones de varios clasificadores en las predicciones finales de cada sistema también ha contribuido a compensar el *overfitting* o *underfitting* que pueda tener algún modelo en concreto.
- Para conseguir **interpretabilidad** se ha diseñado un **Sistema de Predicción e Interpretación** que ha proporcionado los valores de confianza de las predicciones, las predicciones de cada clasificador por separado, las predicciones combinadas, y los mapas de activación.

1.4. Estructura

El presente documento está dividido en los siguientes 7 capítulos:

1. **Introducción:** Este primer capítulo se presenta el problema y la forma en la que éste será abordado en los sucesivos capítulos.
2. **El ojo y sus patologías:** Durante este segundo capítulo se estudia la anatomía y fisiología del ojo y se analizan las características principales las dos patologías que han motivado esta investigación: RD y DMAE.
3. **Machine Learning y aplicaciones médicas:** Además de ofrecer una visión general del funcionamiento y características de los sistemas de Machine Learning, durante estas páginas se muestran ejemplos de las aplicaciones médicas de los mismos.
4. **Estado del arte en detección de RD y DMAE:** Se analizan las principales aproximaciones para la detección de RD y DMAE, tanto de Machine Learning como de Deep Learning, publicadas hasta el momento.
5. **Diseño de Sistema de Detección de RD y DMAE:** En este capítulo se muestra el sistema propuesto para la detección de RD y DMAE. También se detallan las características de todos los conjuntos de imágenes utilizados para el entrenamiento del sistema y el sistema adicional para la interpretación de las predicciones.

6. **Análisis de los resultados obtenidos:** Este capítulo detalla las evaluaciones realizadas al sistema presentado en el capítulo anterior.
7. **Conclusiones:** Para finalizar, se analizan las aportaciones realizadas por esta investigación, su aplicabilidad en el mundo real y las posibles líneas de investigación futuras que se abren en este momento.

Capítulo 2

El ojo y sus patologías

Aunque comúnmente se suele hablar de los ojos como nuestra ventana al exterior (Zhu et al. 2001), la realidad es que su funcionamiento y estructura es considerablemente más complicado que el de una simple ventana de cristal. Dada su extrema perfección, incluso Charles Darwin reconoció tener grandes dificultades para explicar los ojos únicamente mediante variación y selección. (Darwin 2004)

2.1. Anatomía y fisiología ocular

Los ojos son el principal órgano de la visión. La perfección del ojo es tal, que cada ojo ha evolucionado adaptándose a las necesidades del organismo poseedor, lo que ha provocado que existan diversas diferencias en la anatomía y fisiología ocular de los diferentes organismos. (Zhu et al. 2001).

La estructura más simple de ojo consiste en una concentración de células fotorreceptoras mediante las cuales un organismo puede distinguir, no sólo la luz y la oscuridad, sino también la dirección de la luz incidente. Esta última característica supondría, para los organismos con este tipo de sistema ocular, una ventaja evolutiva ante otros tipos de organismos que únicamente podrían diferenciar entre luz y oscuridad.

Sin embargo, el sistema óptico complejo presente en el 96 % de las especies

animales, es capaz de realizar un proceso completo que comienza con la detección de la luz y finaliza con unos impulsos electroquímicos viajando a través de las neuronas. Durante ese proceso, los ojos tienen que captar la luz, regular la intensidad mediante un diafragma y, mediante un sistema de lentes (cristalino), enfocarla en único punto que se encargará de realizar la transformación en impulsos eléctricos. Este punto donde convergen todos los rayos de luz, que será objeto de estudio durante este trabajo, es conocido como **retina**.

La anatomía y fisiología ocular (Figura 2.1) es similar en la mayoría de los vertebrados. El globo ocular, que contiene el resto de elementos del sistema, es una esfera llena de **humor acuoso**, que es un líquido compuesto en un 99 % por agua. El constante flujo de este líquido en el ojo permite regular la presión ocular, de forma que las propiedades del ojo puedan mantenerse constantes. Además, también permite aportar nutrientes y oxígeno a la parte anterior del ojo y eliminar deshechos de esta zona a la que los capilares no son capaces de llegar (Zhu et al. 2001).

La pared del globo ocular la forman 3 capas conocidas como (desde la más interna a la más externa): **retina**, **coroides** y **esclerótica**. Cuando la luz llega al ojo, el primer elemento con el que tiene contacto es la **córnea**, que pertenece a la capa esclerótica. Debido a su índice de refracción (mayor que el del aire), la córnea provocará que se desvíen los rayos de luz que lleguen a ella permitiendo, así, que converjan en el centro del ojo. La cornea protege al resto del ojo de polvo, gérmenes y cualquier tipo de sustancia dañina. Además, también filtra los rayos ultravioleta procedentes de la luz solar (Zhu et al. 2001). La mínima dispersión que se produce en los rayos de luz, que nos permite obtener una imagen clara y definida, está asegurada por la uniformidad espacial de sus células (Oyster 1999).

Posteriormente, es el **iris** quien se encargará de contraer o expandir la **pupila**, lo que permitirá regular la cantidad de luz que entra al ojo. Esta es la razón por la que, en condiciones de baja luminosidad, nuestras pupilas se ven dilatadas, para poder permitir el paso de la mayor cantidad posible de luz.

A continuación, y como en otros sistemas ópticos artificiales, necesitamos un

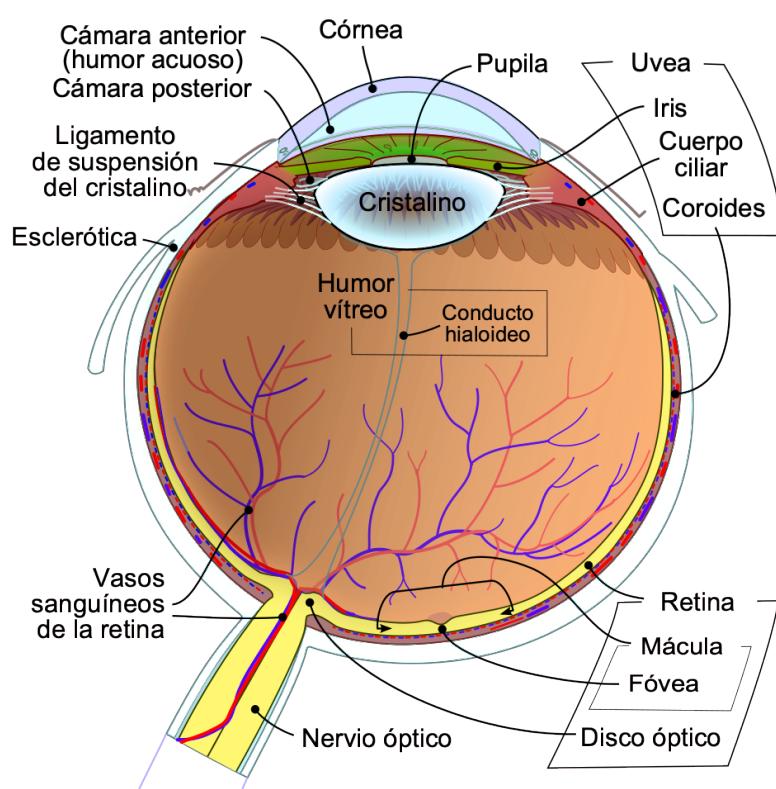


Figura 2.1: Estructura del ojo humano. Fuente: Wikipedia

elemento que enfoque toda esa luz en un único punto. Este proceso se realiza mediante el **cristalino** que actúa como lente, y una serie de músculos a su alrededor que modifican su forma (y su índice de refracción) para permitirnos enfocar objetos a diferentes distancias. Al igual que pasaba con la córnea, es necesario que los elementos que forman el cristalino tengan un índice de refracción mayor que el de la córnea y del humor acuoso, lo que le permitirá enfocar correctamente.

Una vez atravesado el cristalino, la luz llegará a la **retina** donde se producirá la transformación de la luz en impulsos eléctricos, que posteriormente viajarán por el nervio óptico hasta el cerebro, que será capaz de procesar y comprender la imagen recibida.

2.1.1. LA RETINA Y SU IMPORTANCIA

La palabra **retina** procede del latín medieval **rete** o **retis** (red). Toma ese nombre debido a la gran red de vasos sanguíneos que la forman. Utilizando términos de ingeniería, la retina es el transductor en el proceso de visión. Es la capa de tejido sensible a la luz situada en el fondo del ojo sin la cual todo el proceso detallado anteriormente carecería por completo de sentido, puesto que el cerebro no recibiría la información captada por los ojos. Su color, rojo, es debido a la inmensa cantidad de vasos sanguíneos que existen detrás de ella.

A nivel macroscópico, la retina está formada por los siguientes elementos (Figura 2.2):

- **Papila o disco óptico:** Conocido como *punto ciego* debido a la ausencia de fotoreceptores, es el punto de entrada del nervio óptico en el globo ocular. Tiene un diámetro aproximado de 1.5 mm y forma circular de color amarillo. A través del disco óptico entra al globo ocular la arteria central de la retina y sale la vena central de la retina. En el disco óptico encontramos también una excavación fisiológica conocida como **cúpula o copa**. Su tamaño, y más concretamente, el cociente entre su diámetro y el del disco óptico es un buen indicador para la

detección de la enfermedad conocida como glaucoma.

- **Arterias y venas:** Son las encargadas de proveer de oxígeno y nutrientes a la retina. La arteria central de la retina entra en el ojo a través del nervio óptico y se separa en dos ramas, que a su vez se separarán formando una extensa red de capilares. Muchas de las enfermedades de la vista afectan a estos vasos sanguíneos, bloqueándolas o haciéndolas más frágiles.
- **Mácula:** Esta pequeña área con gran pigmentación se encuentra en el centro de nuestra retina. La mácula tiene un diámetro aproximado de 5 mm. Es la encargada tanto la visión central como de la visión en detalle y en movimiento.
- **Fóvea:** Es una hendidura en el centro de la mácula, con un diámetro aproximado de 1.0 mm que permite enfocar los rayos que llegan a la retina.
- **Retina periférica:** Como su nombre indica, nos permite la visión periférica, es decir, la de los rayos de luz que no están en nuestro foco central de visión.

A nivel microscópico, la retina tiene una estructura compleja formada por varias capas de neuronas interconectadas. Existen dos tipos principales de fotoreceptores en la retina: los **conos** y los **bastones**. Las células de la retina presentan grandes similitudes con las del cerebro, apoyando la afirmación común de que el sistema visual es una extensión del sistema nervioso central (Zhu et al. 2001).

Estos receptores contienen unos productos químicos conocidos como **foto-pigmentos**. Los fotopigmentos tienen la propiedad de descomponerse ante la exposición a la luz, excitando en el proceso a las fibras nerviosas que salen del ojo.

Los **bastones** son estructuras cilíndricas y alargadas extremadamente sensibles a los cambios de intensidad de la luz. Sin embargo, no son capaces de percibir información sobre el color. De esto se encargan los **conos**, que son células más pequeñas y finas capaces de percibir el color y de capturar detalles más finos.

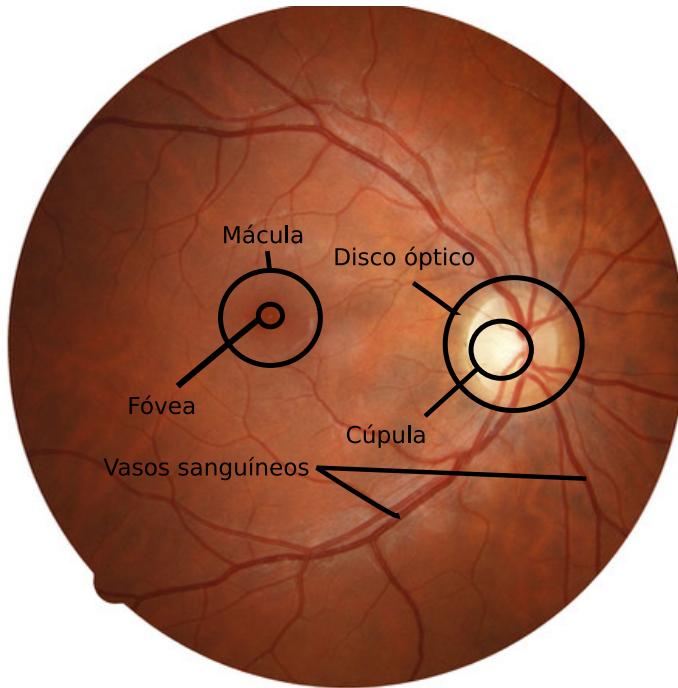


Figura 2.2: Elementos de la retina. Fuente: Kaggle (anotaciones de elaboración propia)

En los **conos** encontramos tres tipos distintos de fotopigmentos que responden a diferentes longitudes de onda distinta de la luz. Esto da lugar a los conocidos como colores primarios de la luz: rojo, azul, y verde.

En la **fóvea central**, los únicos fotoreceptores existentes son los conos, encargados de la visión en detalle y visión en color. Según nos alejamos de la fóvea y nos dirigimos hacia la parte más periférica de la retina, los bastones empiezan a ser predominantes. Estos son responsables de la visión periférica y la visión en bajas condiciones de luminosidad.

En la retina humana existen aproximadamente 125 millones de fotoreceptores, de los cuales, aproximadamente 120 millones son bastones y 5 millones son conos.

Conectadas a los conos y bastones encontramos las **células ganglionares**, un tipo de neuronas en la superficie interna de la retina en las que se produce una diferencia de potencial que se transmite a través de su largo axón hasta el tálamo, hipotálamo y mesencéfalo del cerebro. Ya en el cerebro, esta información es procesada e interpretada por el **córtex visual**.

2.2. Principales patologías de la retina

Existen dos tipos principales de enfermedades que afectan a la retina: las enfermedades vasculares y las degenerativas. Durante este trabajo analizaremos dos de las más importantes: la **Retinopatía Diabética (RD)** y la **Degeneración Macular Asociada a la Edad (DMAE)**. En la Figura 2.3 podemos ver el efecto que tienen estas en la visión.

Aún siendo de naturaleza distinta y provocando distintos efectos, ambas patologías tienen algo en común: **la mayoría de casos de ceguera provocados por ellas hubieran podido ser evitados con una detección y tratamiento de las mismas en los primeros estadios**. La detección de estas, como veremos más adelante, pasa comúnmente por el análisis de la retina mediante imágenes de fondo de ojo. Este tipo de imágenes permiten proyectar la estructura 3D de la retina en un plano 2D. Para captarlas utilizamos cámaras de fondo de ojo, un tipo especial de cámaras que cuentan con un microscopio de baja potencia con una cámara adherida, permitiendo una factor de magnificación de 2.5x. Los rayos de luz viajan desde la retina a la cámara atravesando la pupila. El sensor de la cámara es un sensor RGB similar al de otros tipos de cámaras.

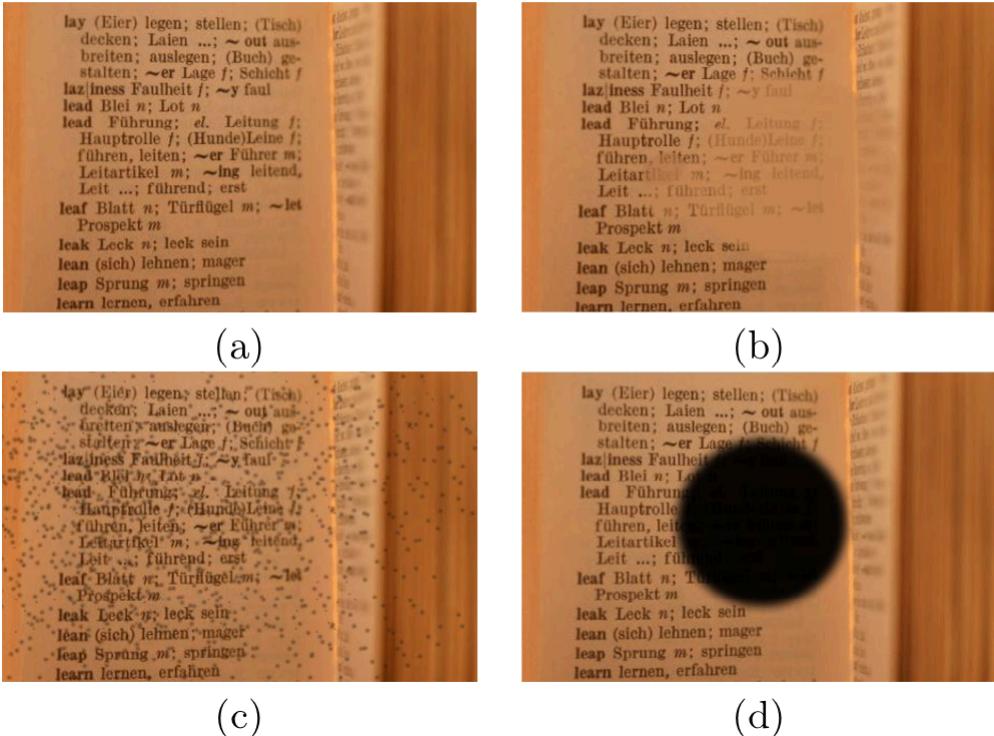


Figura 2.3: Efectos en la visión de las enfermedades analizadas en este trabajo: (a) Visión normal, (b) con Retinopatía Diabética no proliferativa, (c) con Retinopatía Diabética proliferativa y (d) con degeneración macular asociada a la edad. Fuente: American Academy of Ophthalmology (www.aao.org)

2.2.1. RETINOPATÍA DIABÉTICA

Las personas que sufren de diabetes presentan altos niveles de azúcar en sangre debido a la incapacidad de su páncreas de generar suficiente insulina para distribuir el azúcar (Diabetes Tipo I) o a la incapacidad del organismo de asimilar correctamente la insulina (Diabetes Tipo II). Estos altos niveles de azúcar pueden producir daños en varios organismos presentes en nuestro cuerpo.

La Retinopatía Diabética ocurre cuando, debido a la diabetes, se dañan los vasos sanguíneos de la retina. Es común establecer dos etapas principales de RD: proliferativa y no proliferativa.

- **RD No Proliferativa (NPDR):** Es el primer estadio de la enfermedad. Durante esta etapa aparecen microaneurismas, pequeñas áreas de inflamación en los vasos sanguíneos de la retina. Además, algunos vasos sanguíneos se obstruyen. En casos más complicados, el bloqueo de una gran cantidad de vasos sanguíneos provoca que haya áreas de la retina que dejen de recibir sangre por completo.
- **RD Proliferativa (PDR):** En esta etapa, de mayor gravedad, las áreas de la retina que no estaban recibiendo sangre, envían señales al cuerpo para que se hagan crecer nuevos vasos sanguíneos. Sin embargo, estos nuevos vasos sanguíneos son frágiles y anormales, y en el caso de rotura y goteo de sangre, podrían provocar una pérdida severa en la visión o incluso resultar en ceguera total.

La detección temprana de la RD ha demostrado ser de vital importancia para evitar la pérdida de vista e incluso la ceguera causada por la misma. Los primeros estadios de la Retinopatía Diabética son casi asintomáticos, y no empiezan a afectar a la visión del paciente hasta que la enfermedad ha avanzado a un estadio en el que el tratamiento es mucho más complicado y costoso. Se recomienda a los pacientes diabéticos al menos un análisis anual, para poder aplicar un tratamiento de la Retinopatía Diabética a tiempo (Fong et al. 2004).

El tratamiento de la Retinopatía Diabética más común es la **fotocoagulación**.

ción con láser. Este tratamiento se puede realizar en una o varias sesiones, tras haber comprobado, mediante una angiografía fluoresceínica el estado de los vasos sanguíneos. Además, este tratamiento puede ir acompañado de inyecciones intravítreas de medicación antangiogénica, que se encargará de evitar el desarrollo excesivo y anormal de los vasos sanguíneos. En casos de gravedad, puede ser preciso recurrir a la **vitrectomía**, una técnica de microcirugía intraocular.

Las lesiones típicas derivadas de la Retinopatía Diabética son:

- **Exudados duros:** Son depósitos lipídicos de color amarillento brillante y bien definidos, que se filtran procedentes de los vasos sanguíneos de la retina. Suelen encontrarse en la capa más externa de la misma (Group & others 1991).
- **Exudados blandos (o manchas algodonosas):** Son engrosamientos isquémicos de la capa de fibras nerviosas. Presentan bordes difusos y un color blanco.
- **Microaneurismas:** Aparecen normalmente como pequeños grupos de puntos rojos con bordes muy definidos. Son causados por la dilatación de pequeñas venas, y son uno de los primeros signos de Retinopatía Diabética no proliferativa (Williams et al. 2004). Los microaneurismas suelen tener bordes bien definidos y su tamaño suele variar entre los $20\mu\text{m}$ y $200\mu\text{m}$, lo que supone menos de un 8 % del tamaño total del disco óptico (Group & others 1991).
- **Hemorragias:** Son pequeñas manchas rojas con diversas formas y márgenes ligeramente definidos que aparecen en las imágenes de fondo de ojo debido a los puntos de sangrado en la retina. Suelen tener un tamaño de unos $125\mu\text{m}$ (Group & others 1991).

En la Figura 2.4 se observan algunas de las lesiones descritas anteriormente.

Además, como hemos visto anteriormente, en la RD Proliferativa se produce la neovascularización, aparición de nuevos vasos sanguíneos en la retina (Figura 2.5).

La tabla 2.1 nos muestra una posible clasificación de los diferentes estadios de

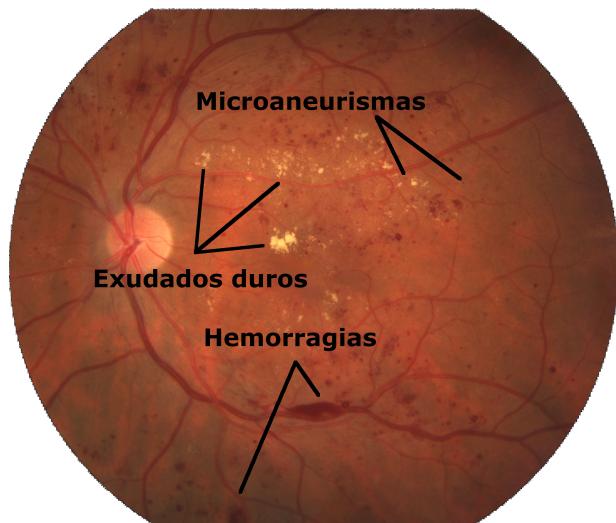


Figura 2.4: Lesiones típicas de la Retinopatía Diabética. Elaboración proia

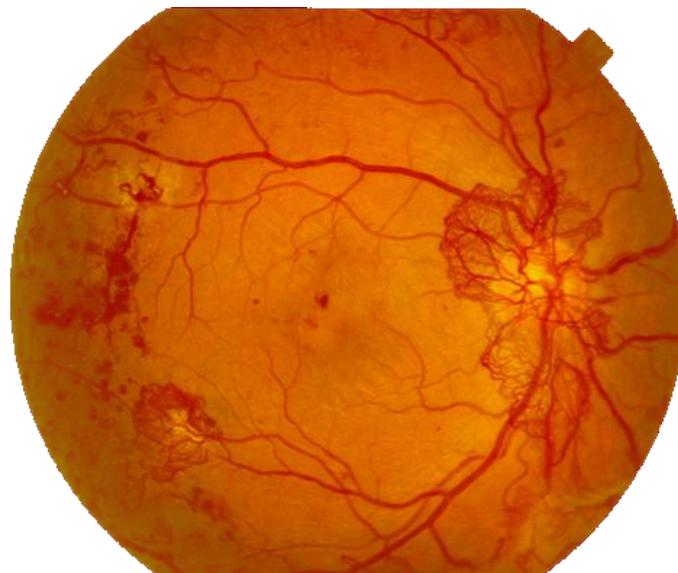


Figura 2.5: Ejemplo de retina en la que se ha producido neovascularización

la Retinopatía Diabética en función de las lesiones presentes en el paciente.¹ Los pacientes con Retinopatía Diabética No Proliferativa en grado 3, tienen un 50 % de probabilidad de desarrollar Retinopatía Diabética Proliferativa en menos de un año (Mansour 2017).

Tabla 2.1: Niveles de gravedad de la Retinopatía Diabética en función de las lesiones observadas

Nivel de gravedad	Observaciones
Grado 0: No RD	Sin ninguna anomalía
Grado 1: NPDR Ligera	Presencia de algunos microaneurismas
Grado 2: NPDR Moderada	Presencia de más microaneurismas pero menos que en el grado 3
Grado 3: NPDR Severa	Alguno de los siguientes: - Más de 20 hemorragias intrarretinales - Dilataciones venosas - Anomalías microvasculares en la retina - Ningún signo de PDR
Grado 4: PDR	Alguno de los siguientes: - Neovascularización - Hemorragia vítrea

Existen, incluso, estudios que han demostrado que los pacientes que padecen de Retinopatía Diabética Proliferativa, sufren más riesgo de tener ataques al corazón, amputaciones o nefropatía diabética (Klein et al. 1984), (Cade 2008), (Acharya et al. 2009).

2.2.2. DEGENERACIÓN MACULAR ASOCIADA A LA EDAD

La Degeneración Macular Asociada a la Edad (DMAE) afecta a la mácula provocando que, quien la sufre, comience a ver imágenes desenfocadas o deformadas y con zonas oscurecidas. Como se ha explicado previamente, la mácula permite la visión central, y su degeneración afecta directamente al día

¹Basada en la clasificación de <https://idrid.grand-challenge.org/grading/>

a día del paciente incapacitándolo para hacer tareas comunes como pueden ser la lectura o la conducción.

Aunque en sus primeros estadios la progresión de la enfermedad sea muy lenta y el paciente puede que únicamente perciba un ligero cambio en su visión, en fases avanzadas la DMAE puede provocar la pérdida total de la visión central. Si es detectada a tiempo, la DMAE puede ser retardada y mitigada mediante vitaminas y minerales.

El principal signo de DMAE en las imágenes de fondo de ojo es la aparición de las **drusas**, depósitos amarillos localizados bajo la retina, procedentes de la acumulación de minerales. En la Figura 2.6 se observa la forma de las drusas. En función del número y tamaño de drusas, podemos definir tres estadios en la enfermedad:

- **Estadio inicial:** En este estadio existe un reducido número de drusas redondas y de pequeño tamaño (menos de 125 μm). Además, estas tienen unos bordes bien definidos. Los pacientes con este grado de DMAE no sufren pérdida de visión y tienen un riesgo bajo de desarrollar complicaciones (Group & others 2001).
- **Estadio intermedio:** En este estadio existen muchas más drusas, y estas tienen un tamaño mayor, llegando incluso a aparecer algunas de más de 125 μm . Se pueden apreciar cambios en la pigmentación de la retina. Cuando las drusas dejan de tener bordes definidos, y aparecen en grupos, el riesgo de complicaciones de la DMAE es mucho mayor.
- **Estadio avanzado:** Existen dos subniveles dentro del estadio avanzado:
 - **DMAE seca o atrófica:** Se produce por la acumulación de desechos que atrofian las células fotosensibles de la zona macular. Es la forma más común, y tiene una evolución lenta y progresiva.
 - **DMAE húmeda o exudativa:** En la DMAE húmeda, crece una membrana vascular bajo la retina. De la misma forma que en la Retinopatía Diabética Proliferativa, estos nuevos vasos sanguíneos son muy frágiles y pueden romperse derramando líquido, lo que afectará severamente a la visión.



Figura 2.6: Retina con drusas a causa de DMAE

2.2.3. SISTEMAS DE DIAGNÓSTICO

Uno de los grandes problemas de la Retinopatía Diabética es que no existe ninguna señal que nos avise en estadios muy tempranos de la enfermedad, y en el momento que los usuarios deciden hacerse una examinación, suele ser demasiado tarde para un tratamiento óptimo.

Existen varios tipos de sistemas para el diagnóstico de la RD y DMAE, entre los que destacan las **fotografías de fondo de ojo**, la **tomografía de coherencia óptica (OCT)** o la **angiografía**. La fotografía de fondo de ojo, a diferencia de las otras dos técnicas, puede ser realizada con sistemas relativamente baratos y fáciles de manejar y de transportar. Además, las cámaras de fondo de ojo pueden capturar la información de la retina mediante técnicas no invasivas. En función de la patología que se intente diagnosticar, estas imágenes están centradas en la mácula o en el disco óptico. La Figura 2.2 analizada anteriormente para explicar las partes de la retina, es un ejemplo de las imágenes que proporcionan este tipo de cámaras, con tamaños de hasta 16 megapíxeles. Es por ello por lo que la fotografía de fondo de ojo es el sistema más utilizado en los centros de atención primaria y en el que enfocaremos nuestro estudio. Sin embargo, cabe destacar que en ocasiones

éstas no son suficientes, y tendrán que ser combinadas con otros tipos de sistemas.

El diagnóstico de la Retinopatía Diabética es tradicionalmente realizado por oftalmólogos que inspeccionan las imágenes de fondo de ojo en busca de las diferentes lesiones que caracterizan estas patologías. Sin embargo, este es un proceso que requiere una gran cantidad de tiempo. La limitada cantidad de profesionales capaces de realizar este proceso hace imposible cubrir la demanda actual, que no hace más que crecer (Bjørvig et al. 2002). Este hecho es más acusado en zonas rurales o países no desarrollados donde no es posible el acceso a este tipo de profesionales. El 75 % de los pacientes de Retinopatía Diabética viven en áreas donde no existen especialistas ni infraestructura para la detección y tratamiento de la enfermedad (Guariguata et al. 2014). Por ello, diversidad de **herramientas para el análisis automático de retina (ARIA)** están siendo desarrolladas actualmente. En este tipo de herramientas nos centraremos durante el desarrollo de este trabajo. Éstas llevarán el diagnóstico a sitios donde no sería posible de otra forma, además de reducir costes y reducir el tiempo necesario para los diagnósticos en los lugares donde ya se realizaba de forma manual por un profesional.

Además, según se ha puesto de manifiesto en previos estudios, los profesionales difieren en numerosas ocasiones en el diagnóstico de los diferentes estados de este tipo de patologías, debido a que existe un cierto grado de subjetividad (Ruamviboon et al. 2005), (Sellahewa et al. 2014).

Las técnicas ARIA se basaron en los últimos años en la **extracción de manual de características** de las imágenes de fondo de ojo, que posteriormente se le pasarían a un **clasificador de Machine Learning**. Era, por lo tanto, un proceso con dos fases claramente diferenciadas: **extracción de características y clasificación**. Este proceso requería conocimiento experto para la definición de las características que nos fueran de mayor utilidad para la detección de la RD y la DMAE.

Con la entrada del **Deep Learning**, y concretamente las **Redes Neuronales Convolucionales (CNN)** este proceso inicial de extracción de características ha podido ser automatizado, mejorando la calidad de los sistemas notablemente. Las fases de extracción de características y de predicción se

unifican. Las características extraídas poseen un poder mucho mayor de predicción, puesto que toda la red ha sido entrenada para ello.

Sin embargo, la naturaleza del problema, la **falta de estandarización** y la **escasa cantidad de imágenes etiquetadas**, han provocado que estos sistemas tuvieran serias dificultades para su aplicación general.

Esta aproximación al problema del diagnóstico de la RD y DMAE plantea una serie de preguntas. ¿Cómo se incorporaría un sistema de este tipo en las consultas? ¿Se introduciría el software en las propias cámaras que captan la imagen de la retina? ¿Sería posible crear en los países un sistema centralizado de diagnóstico con imágenes de fondo de ojo? ¿Podrían utilizarse en este sistema centralizado técnicas ARIA combinadas con las opiniones de expertos? No hay una respuesta para todas estas preguntas y las ventajas e inconvenientes de algunas de ellas se desarrollarán en capítulos posteriores.

Capítulo 3

Machine Learning y aplicaciones médicas

Tradicionalmente, el trabajo de los ingenieros de software ha consistido en dar a las computadoras una serie de reglas explícitas de cómo tienen que procesar la información para tomar decisiones. Sin embargo, la complejidad del campo de la medicina es tal que sería prácticamente imposible capturar toda la información relevante mediante una serie de reglas definidas de forma explícita (Schwartz et al. 1986).

El **Machine Learning** es la rama de la Inteligencia Artificial que ha permitido crear **sistemas que aprendan de los datos sin necesidad de que se programen reglas específicas**. Esto ha supuesto una auténtica revolución en prácticamente cualquier sector profesional entre los que, por supuesto, se encuentra también la medicina. Estos sistemas buscan, de forma automática, **patrones** en los datos que les permitan predecir una variable objetivo en función de una serie de variables de entrada del sistema. De esta forma se crea un **modelo** que, idealmente, será capaz de generalizar y obtener la salida correcta para nuevas entradas nunca vistas. Esto se conoce como **Aprendizaje Supervisado** aunque es importante mencionar que no es la única forma de Machine Learning o Aprendizaje Automático.¹

¹Existe también, por ejemplo, el Aprendizaje No Supervisado, que permite encontrar patrones en los datos aunque no exista una variable objetivo a predecir

Uno de los principales inconvenientes del Machine Learning con respecto al aprendizaje humano es, a la vez, una de sus principales ventaja: la **necesidad de grandes cantidades de datos para su correcto funcionamiento**. Si se alimentan con una cantidad suficiente de datos, los algoritmos de Machine Learning podrán encontrar patrones que, para los humanos, serían prácticamente imposible de detectar. El cerebro humano es una máquina bastante compleja y sofisticada de encontrar patrones. Sin embargo, tiene grandes dificultades en realizar el análisis de datos con alta dimensionalidad. Un modelo de Machine Learning podrá analizar, en segundos, más pacientes de los que verá un médico en toda su vida. Además, la cantidad de predictores distintos que manejará sería totalmente inviable para un humano.

En la Figura 3.1 vemos como, a pesar de existir desde los años 60, el interés de la población en el Machine Learning ha experimentado un gran ascenso en los últimos años. La democratización del Machine Learning ha comenzado y multitud de empresas han empezado a usar modelos predictivos de Aprendizaje Automático en sus procesos. Existen 3 principales motivos en este crecimiento:

- **Nuevos algoritmos:** Principalmente en la rama del Deep Learning, en los últimos años se ha producido una serie de importantes avances. Sin embargo, este no es el factor principal del crecimiento, pues la mayoría de algoritmos que se están implantando en muchas compañías existen desde hace varias décadas.
- **Mayor capacidad de computación:** Sin duda, este ha sido un factor clave en el crecimiento de estas técnicas. Además, la entrada al mercado de las tarjetas gráficas o GPUs ha permitido paralelizar los procesos consiguiendo ejecuciones cientos de veces más rápidas.
- **Mayor cantidad de datos:** Todos estos algoritmos no podrían aportar valor de no existir ingentes cantidades de datos, tanto estructurados como no estructurados, en los que poder encontrar patrones. Con el creciente uso de servicios online y la expansión del IoT o Internet de las Cosas se están generando mayor cantidad de datos cada día que nunca antes se había generado. Según Forbes, el 90 % de los datos existentes

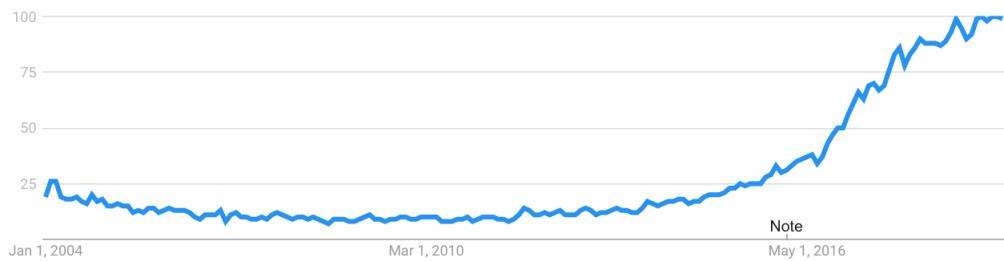


Figura 3.1: Interés, a lo largo del tiempo y en todo el mundo, del término Machine Learning en el buscador Google. Datos de Enero de 2014 a Julio de 2019. Un valor de 100 indica la máxima popularidad del término. Los valores 50 y 0 indican, respectivamente, que un término es la mitad de popular en relación con el valor máximo o que no existen suficientes datos del término. Fuente de los datos: Google Trends

en 2018 en todo el mundo, se generaron entre 2016 y 2017.²

3.1. IA, Big Data, Machine Learning y Deep Learning

Inteligencia Artificial, Big Data, Machine Learning, Deep Learning; actualmente existe mucha confusión en el uso de estos términos. Aunque comparten características, no tienen el mismo significado. En este apartado se detallarán las similitudes y diferencias entre todos ellos para evitar el lenguaje inexacto usado habitualmente, principalmente, en publicidad y medios de comunicación.

Comenzaremos por el **Big Data**, pues es el término más vago y confuso. Cuando hablamos de Big Data nos referimos al análisis de grandes cantidades de datos que no podrían ser analizados con técnicas convencionales de computación. Sin embargo, las líneas que marcan las fronteras del Big Data están difusas, y a menudo es un término más utilizado por medios de comunicación y falsos gurús que por profesionales técnicos y académicos.

Por otro lado, los campos de la **Inteligencia Artificial (IA)**, el **Machine Learning** y el **Deep Learning** sí que están más claramente definidos aunque, el hecho de que cada uno de ellos sea un subcampo del anterior (Figura

²<https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/>

3.2), a menudo da lugar a confusión. Llamamos **Inteligencia Artificial** a un conjunto de técnicas que tratan de que los ordenadores imiten, de alguna forma, el comportamiento humano.

El **Machine Learning** es un subcampo dentro de la IA, que consiste en un conjunto de técnicas y herramientas, principalmente estadísticas, que permiten a los ordenadores obtener patrones a partir de grandes conjuntos de datos. Gracias a esos patrones seremos capaces de entender mejor los datos o hacer predicciones. La forma más común de Machine Learning es el conocido como **Aprendizaje Supervisado**. Durante el entrenamiento de los modelos de Aprendizaje Supervisado, se proporcionan al algoritmo una serie de datos históricos. Entre ellos se encuentra la **variable objetivo**, es decir, la que posteriormente querremos predecir en los nuevos datos de entrada. Por ejemplo, en un modelo de detección de cáncer a partir de imágenes médicas, nuestra variable objetivo será precisamente la que indique si una imagen pertenece a un paciente enfermo de cáncer o un paciente sano. Esta variable, por lo tanto, tendrá dos posibles valores, siendo este un problema de **clasificación**. En los problemas de clasificación se tratan de predecir **variables discretas o clases**, es decir, variables que solo pueden tomar un rango limitado de posibles valores. Si, por ejemplo, realizáramos un modelo para predecir el precio de una vivienda en función de sus características, nos encontraríamos ante un problema de **regresión**, pues el precio es un valor continuo.

Es común en los algoritmos para aprendizaje supervisado el uso de una **función de coste**. Esta función mide el error entre las predicciones del modelo y los datos reales. De forma iterativa, muchos de los algoritmos de Aprendizaje Automático tratarán de ajustar una serie de parámetros (o pesos) intentando minimizar esta función. Un claro ejemplo de algoritmos con este comportamiento son las conocidas como **redes neuronales**, de las que explicaremos su funcionamiento en el siguiente apartado.

Precisamente las redes neuronales, son las que dan lugar al **Deep Learning**. Cuando añadimos más complejidad a las redes neuronales somos capaces de detectar patrones mucho menos evidentes, además de tratar problemas complejos sin necesidad de un pre-procesamiento manual previo de los datos que

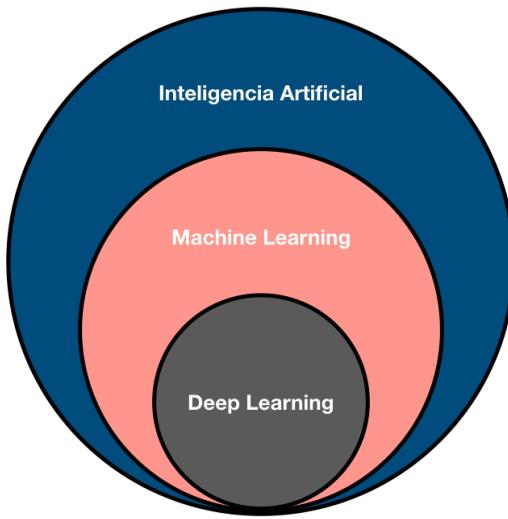


Figura 3.2: El Machine Learning es un campo perteneciente a la Inteligencia Artificial. El Deep Learning, a su vez, es un campo dentro del Machine Learning. Elaboración propia

los simplifique. Este pre-procesamiento sí que es necesario en muchos proyectos de Machine Learning y, de hecho, supone un importante porcentaje del tiempo de trabajo de los ingenieros de Machine Learning. Los algoritmos de Deep Learning son actualmente el estado del arte en tareas como reconocimiento de imágenes (Krizhevsky et al. 2012), reconocimiento del habla (Deng et al. 2013), procesamiento del lenguaje natural (Collobert et al. 2011), análisis de información de aceleradores de partículas (Baldi et al. 2014) o reconstrucción de los circuitos cerebrales (Helmstaedter et al. 2013), entre muchas otras.

Como vemos en la Figura 3.3 el término Big Data, que durante mucho tiempo estuvo en cabeza en popularidad, ha perdido fuerza en los últimos años mientras que Machine Learning y Deep Learning (en menor medida) siguen creciendo.

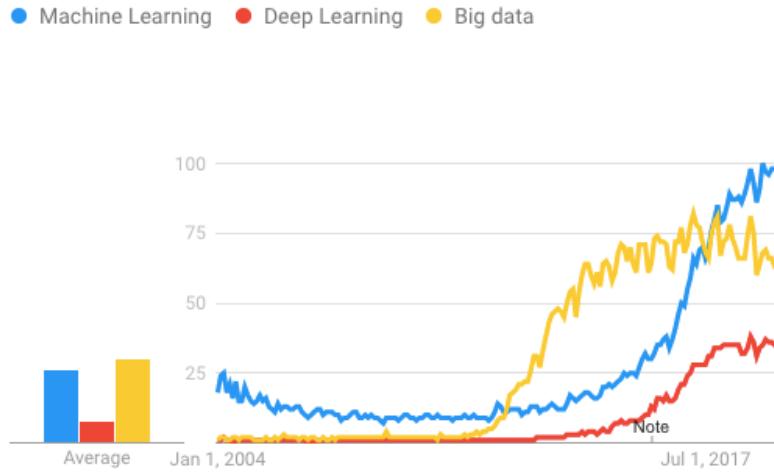


Figura 3.3: Interés, a lo largo del tiempo y en todo el mundo, de los términos Machine Learning (en azul), Deep Learning (en rojo) y Big Data (en amarillo) en el buscador Google. Datos de Enero de 2014 a Julio de 2019. Un valor de 100 indica la máxima popularidad del término. Los valores 50 y 0 indican, respectivamente, que un término es la mitad de popular en relación con el valor máximo o que no existen suficientes datos del término. Fuente de los datos: Google Trends

3.2. Redes neuronales, descenso de gradiente y back-propagation

Una red neuronal consiste en un conjunto de nodos, conocidos como **neuronas**, conectados entre sí para transmitir señales. Estas neuronas suelen estar dispuestas en una serie de **capas**, en las que, comúnmente, cada neurona de una capa está conectada a todas las neuronas de las capas anteriores. De esta forma, la salida de unas neuronas pasa a ser la entrada de otras (Figura 3.4).

La Figura 3.5 representa las operaciones realizadas por una sola neurona durante la predicción. Estas mismas operaciones son realizadas en todas las neuronas de la red. Cada neurona combina sus entradas con un conjunto de coeficientes o pesos. Las entradas x_1, x_2, x_3 y los pesos w_1, w_2, w_3 son números reales, que pueden ser positivos o negativos.³ El nombre de **peso** se debe a que la función de estos, al multiplicarse por los valores de las entradas, es definir la importancia de cada una de ellas. En cada una de las neuronas,

³Aunque no se haya representado, también existe un término adicional, b (término de sesgo), que no está multiplicado por ningún peso y se suma a z

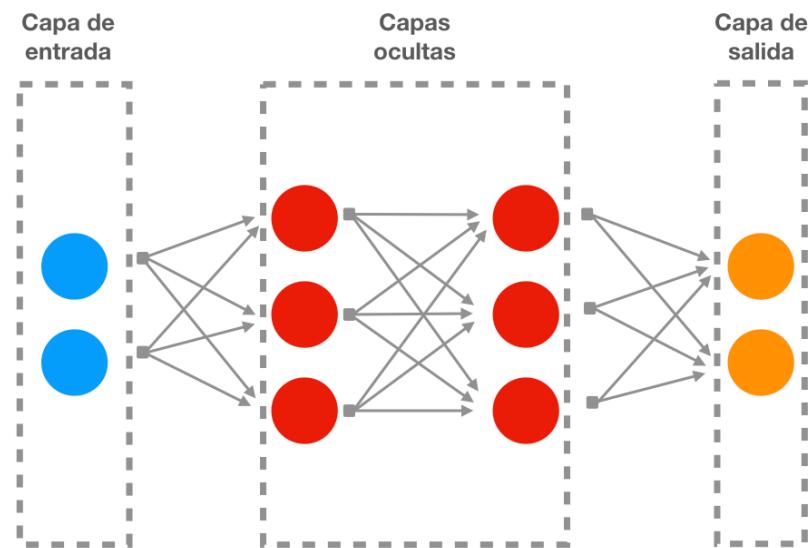


Figura 3.4: Representación de una red neuronal con dos capas ocultas. Cada uno de los círculos representa una neurona. Elaboración propia

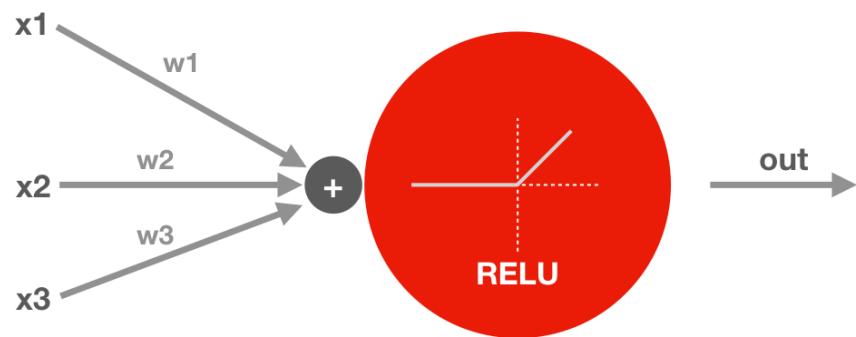


Figura 3.5: Representación de una sola neurona con 3 entradas. Cada una de esas entradas tiene asociado un peso. La neurona utiliza la función de activación ReLU. Elaboración propia

los resultados de todos estos productos se suman (ecuacion 3.2) y se pasa el valor obtenido a lo que se conoce como **función de activación** (ecuacion 3.3) , que añade un comportamiento no-lineal al proceso que permite modelar funciones curvas o no triviales. Actualmente, la función de activación más utilizada es la **ReLU** (Rectified Linear Unit)⁴ cuya fórmula podemos ver en la ecuación 3.1.

$$f(z) = \max(z, 0) \quad (3.1)$$

$$z = x_1w_1 + x_2w_2 + x_3w_3 + b \quad (3.2)$$

$$out = \max(z, 0) \quad (3.3)$$

Como vemos en la ecuación 3.4, las ecuaciones anteriores pueden ser generalizadas para cualquier número de entradas.

$$out(X) = \max\left(\sum_i x_i w_i + b, 0\right) \quad (3.4)$$

Durante el **entrenamiento**, los pesos cambian de valor, intentando minimizar la función de coste. Suponiendo que y es el valor real de la variable objetivo para un conjunto de entradas X , la función de coste $L(X, y)$ podría ser simplemente la de la ecuación 3.5.

$$L(X, y) = (out(X) - y)^2 \quad (3.5)$$

Para ajustar el vector de pesos se suele calcular el vector de gradiente. Este vector indica, para cada peso, cómo se modificaría el error si ese peso se aumentara ligeramente. Es decir, nos proporciona la pendiente de la función de coste (o función de pérdidas). El vector de pesos es entonces ajustado

⁴Otras funciones de activación usadas comúnmente son **softmax**, **tangente hiperbólica** o la **función sigmoide**

en el sentido opuesto al vector de gradiente (ecuación 3.6), buscando así **minimizar el error**. El valor α representa lo que conocemos como **learning rate** o **factor de aprendizaje** y se encarga de controlar la velocidad a la que la red neuronal aprende. Es muy importante la elección correcta de este parámetro, pues, un valor demasiado bajo supondrá que la red tarde muchas iteraciones en encontrar el mínimo de la función de coste. Sin embargo, un valor demasiado alto puede suponer que la red no sea capaz de converger y encontrar este mínimo. Este proceso completo es lo que conocemos como **descenso de gradiente**.

$$w_{ij} = w_{ij} - \alpha \frac{\partial L(X, y)}{\partial w_{ij}} \quad (3.6)$$

En la práctica, este proceso no usa todos los datos cada vez sino que se utiliza el **Descenso de Gradiente Estocástico** (SGD por sus iniciales en inglés). Gracias al SGD podemos actualizar los pesos de nuestra red neuronal tomando cada vez un pequeño conjunto de datos (conocido como **batch**).

El origen de las redes neuronales es el **Perceptrón**, desarrollado en los años 60, que era una red simple de una sola capa de entrada y una capa de salida. Sin embargo, fue en los años 80 cuando estas comenzaron a desarrollar su verdadero potencial gracias al algoritmo de *backpropagation*, que permitió que se añadieran nuevas capas intermedias a las redes neuronales, conocidas como **capas ocultas**. La técnica de *backpropagation* no es más que una aplicación de la regla de la cadena de las derivadas que permite propagar el error calculado al final de la red a todas las capas de ésta. En la ecuación 3.7 podemos ver un ejemplo de como aplicar la regla de la cadena de las derivadas para obtener la derivada de la función de coste en función de los pesos. De la misma forma, podríamos aplicar la regla de la cadena para obtener la derivada de la función de coste en función de los pesos de varias capas atrás.

$$\frac{\partial L(X, y)}{\partial w_{ij}} = \frac{\partial L(X, y)}{\partial \text{out}(X)} \frac{\partial \text{out}(X)}{\partial w_{ij}} \quad (3.7)$$

Gracias a la técnica de *backpropagation*, podemos propagar el error a lo largo

de las capas, para calcular en cada una el vector de gradiente y actualizar con él los pesos. El *backpropagation* ha permitido, por lo tanto, añadir **nuevas capas intermedias** a las redes.

Estas capas intermedias permiten encontrar patrones más complejos, y dieron lugar a lo que conocemos como **Deep Learning**. Si no tuviéramos **capas ocultas**, nuestras redes únicamente encontrarían relaciones directas entre las entradas y las salidas. Sin embargo, las capas ocultas nos permiten modelar de forma mucho más acertada el mundo real, donde las salidas dependen de las interacciones y combinaciones entre las distintas entradas. Estrictamente hablando, nos referimos a Deep Learning cuando tenemos una red con más de una capa oculta. El Deep Learning permite crear modelos computacionales compuestos de múltiples capas de procesamiento que son capaces de aprender representaciones de los datos con **múltiples capas de abstracción** (LeCun et al. 2015).

En las redes profundas, cada capa de neuronas se entrena, automáticamente, en un conjunto de características distinto en base a la salida de la capa anterior. A medida que avanzamos a través de la red, las características que las neuronas son capaces de detectar son más complejas, ya que agregan y recombinan características de capas anteriores. Esta propiedad, conocida como **jerarquía de características**, hace posible que este tipo de redes sean capaces de tratar datasets de muy alta dimensionalidad. Las redes neuronales profundas realizan por lo tanto **extracción automática de características** sin la necesidad de la intervención de un humano (LeCun et al. 2015).

Otra técnica a destacar, que será usada en los sistemas diseñados, es la técnica del **dropout**. Esta técnica de regularización trata de evitar el sobreajuste de la red ignorando de forma aleatoria, durante el entrenamiento, la salida de algunas neuronas. De esta forma, se fuerza a la red neuronal a encontrar patrones más robustos, evitando así que aprenda el *ruido* de nuestro conjunto de datos.

3.2.1. REDES NEURONALES CONVOLUCIONALES

La capacidad de las redes neuronales de encontrar patrones complejos en datasets con una gran cantidad de dimensiones las convierte en candidatas perfectas para tareas como la clasificación de imágenes o el reconocimiento de voz. Sin embargo, estos clasificadores necesitan un trabajo manual previo de extracción de características cuando tratan con señales (imágenes, audios, etc.).

La aparición de las **Redes Neuronales Convolucionales (CNN por sus siglas en inglés)** permitió eliminar la extracción de características y delegarla en el propio algoritmo de *backpropagation*. De esta forma, es posible usar como entradas de nuestro modelo los *datos en bruto* (píxeles de las imágenes o muestras de audio). Un momento clave para las redes convolucionales fue en 2012, en el **ImageNet Large Scale Visual Recognition Challenge (ILSVRC)**⁵ cuando una solución novedosa basada en CNNs (Krizhevsky et al. 2012) obtuvo, de forma holgada, la primera posición en la competición.

La arquitectura de las redes convolucionales está basada en la organización de la corteza visual del cerebro humano. En él, existen neuronas individuales que responden a estímulos en una región delimitada del campo visual. Este tipo de redes son muy similares a las redes neuronales tradicionales analizadas anteriormente. De la misma forma que éstas, las CNN también están compuestas de neuronas dispuestas en capas y se busca minimizar una función de coste mediante el ajuste de una serie de pesos. Sin embargo, las CNN, al asumir que tendrán imágenes como entradas, pueden realizar tareas más especializadas que evitarán la carga computacional que supondría tratar cada píxel de la imagen como un input más de una red neuronal convencional.

Una de las principales ventajas de las redes neuronales convolucionales con respecto a otras aproximaciones al problema es que las CNN poseen un cierto grado de **invarianza a la distorsión y al desplazamiento**. Esto permite que podamos usar este tipo de redes sin apenas pre-procesamiento de las imágenes.

⁵<http://image-net.org/challenges/LSVRC/>

Las CNN constan de **capas convolucionales** y **capas de reducción (o pooling)** alternadas.

En las **capas de convolución** se aplican una serie de **filtros** a las imágenes (cuyos pesos son parámetros modificados durante el entrenamiento por el algoritmo de *backpropagation*). En ellas se producen también las **transformaciones no lineales (ReLU)**. Cada uno de los filtros se desplazará sobre toda la imagen calculándose, en cada posición, el producto escalar entre la región de la imagen y los valores del filtro. Este proceso, la convolución⁶ de la imagen con el filtro, es el que da nombre a estas capas. Estos filtros hacen de **detectores de características**. Precisamente el desplazamiento de ese filtro por toda la imagen es lo que nos permitirá detectar formas y patrones en cualquier posición de la imagen, consiguiendo así la deseada invarianza al desplazamiento. En la Figura 3.6 podemos ver el efecto de la convolución sobre una imagen.

En las **capas de reducción o pooling** se disminuye la cantidad de parámetros. Para ello, se obtiene el promedio o el máximo de una serie de regiones, reduciendo así el tamaño del mapa de características y contribuyendo a evitar el *overfitting*. En función de si se obtiene el promedio o el máximo de las regiones, estas capas son de **Max Pooling** o de **Average Pooling**. La Figura 3.7 representa este proceso.

Al final de todas estas capas tenemos las **Fully Connected Layers**, capas como las de las redes tradicionales que, a partir de los parámetros extraídos por las capas convolucionales y de pooling, realizan las clasificaciones o regresiones finales.

La Figura 3.8 representa todo este proceso en un ejemplo de reconocimiento de dígitos en imágenes. En ella podemos ver la salida de los filtros de las dos capas convolucionales que tiene la arquitectura del ejemplo.

El funcionamiento del algoritmo de *backpropagation* en las redes convolucionales es prácticamente igual que en las no convolucionales, por lo que no supone demasiada dificultad teórica añadida para el entrenamiento. La red

⁶Aunque es común en la literatura hablar de este proceso como convolución, en realidad este cálculo en tratamiento digital de señal es conocido como una correlación cruzada. (Goodfellow et al. 2016)

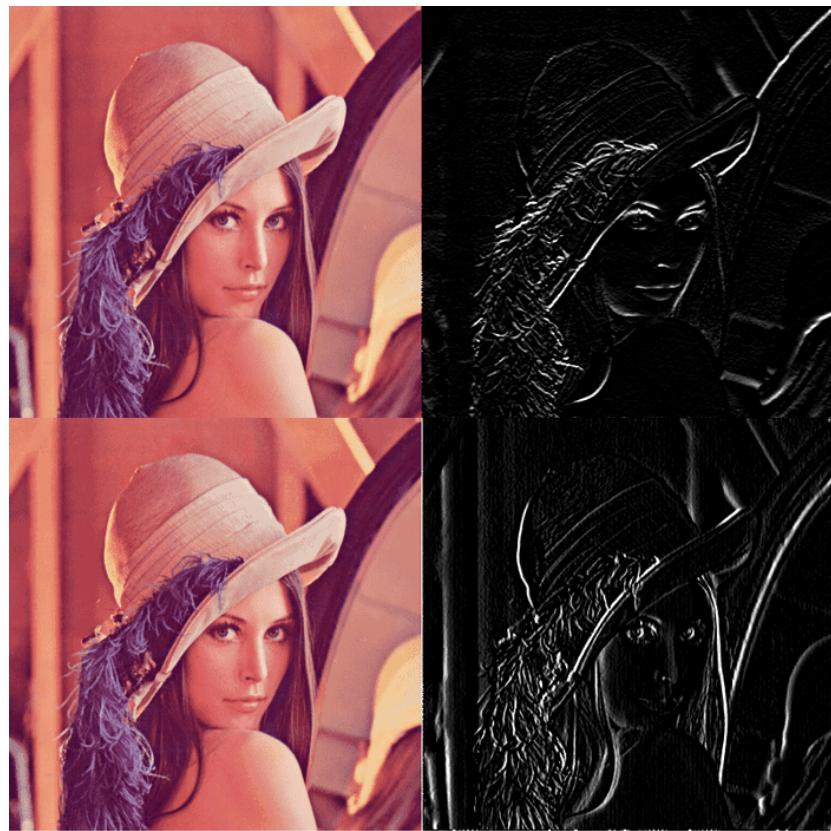


Figura 3.6: Resultado de la convolución de una imagen con un filtro Sobel de 3x3 horizontal (arriba) y otro vertical (abajo) Fuente: <https://victorzhou.com/blog/intro-to-cnns-part-1/>

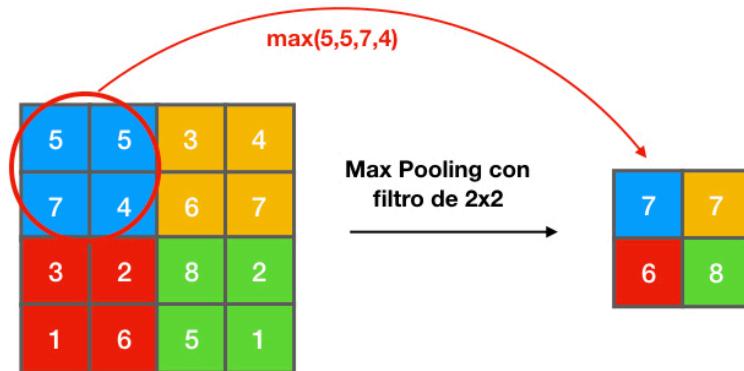


Figura 3.7: Representación del proceso de Max Pooling con un filtro de 2x2 sobre una imagen de 4x4. Elaboración propia

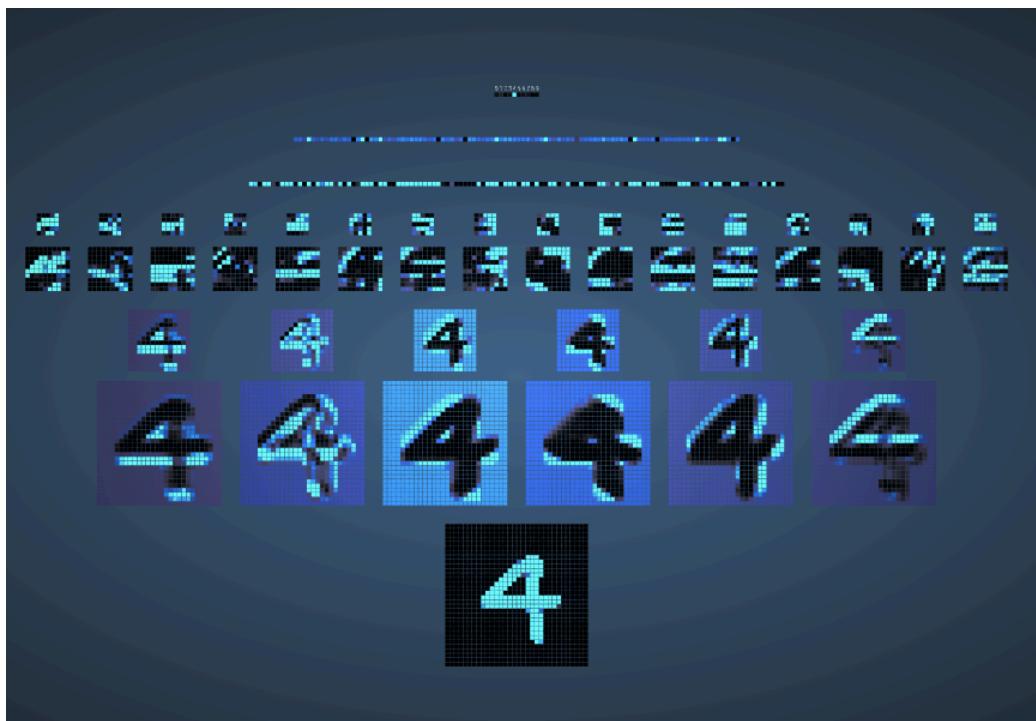


Figura 3.8: Representación de los mapas de activación de una red convolucional con 2 capas convolucionales y 3 fully-connected. Cada capa convolucional va seguida de una de max pooling. Fuente: <http://scs.ryerson.ca/~aharley/vis/conv/flat.html>

será capaz de encontrar, durante el entrenamiento, los pesos de los filtros que permitan extraer las características adecuadas para predecir correctamente nuestra clase objetivo.

Las CNN explotan la propiedad de que, los patrones detectados, no son más que composiciones de otros patrones más simples. En una imagen, por ejemplo, mediante la composición de varias líneas simples damos lugar a motivos que, de nuevo mediante composición, dará lugar a las formas de los objetos. La detección de cada uno de estos niveles de abstracción corresponderá a unas capas concretas de nuestra red convolucional, siendo las primeras capas las que detectarán características más simples como líneas, bordes o colores y las últimas capas las que detectarán elementos compuestos mucho más complejos. Esto es conocido como la **jerarquía de las capas**.

Existe una gran cantidad de arquitecturas de redes convolucionales, que han demostrado ser eficaces en diversos campos. Ejemplos de ellas pueden ser las siguientes:

- **LeNet**: Fue, en 1998, una de las primeras arquitecturas de CNNs (LeCun et al. 1998). Su propósito era, principalmente, el reconocimiento de dígitos en imágenes. Era una red pequeña con 7 capas, siendo dos de ellas convolucionales, otras dos de tipo pooling y el resto fully-connected.
- **Alexnet**: Fue la ganadora en 2012 del concurso ILSVRC (Krizhevsky et al. 2012), con una arquitectura similar a LeNet pero más profunda, con cerca de 60 millones de parámetros y haciendo uso, entre otras novedades, de la función de activación ReLU.
- **VGGNet**: Fue presentada en 2014 (Simonyan & Zisserman 2014), y aún sigue siendo la arquitectura preferida por la comunidad para la extracción de características de imágenes. Fue la primera arquitectura de CNNs realmente profunda (19 capas). Se caracteriza por ser una arquitectura muy uniforme que usa únicamente filtros de 3x3. Sin embargo, era muy costosa de entrenar, pudiendo llegar a tener hasta 140 millones de parámetros.
- **ResNet**: Presentada un año después a la VGGNet (He et al. 2016), se caracteriza por tener saltos entre capas. La salida de la **capa i** puede

ser la entrada de la **capa $i+2$** .

- **Inception:** La primera versión de esta arquitectura (Google Net) (Szegedy et al. 2015) introdujo importantes novedades entre las que destacaba el uso de varios filtros de distintos tamaños en el mismo nivel, cuyas salidas serían concatenadas. El objetivo era crear redes *más anchas* en vez de *más profundas*.

3.3. Transfer Learning

La mayoría de métodos de Machine Learning asumen que los datos de entrenamiento y los de test vienen de la misma distribución y espacio funcional (Pan & Yang 2009). Por ello, cuando esta distribución cambia, debemos volver a entrenar nuestros modelos desde 0, obteniendo datos totalmente nuevos. El **Transfer Learning**, sin embargo, mediante la transferencia de conocimiento entre modelos, permite transferir información de un modelo entrenado previamente a un modelo nuevo que está siendo entrenado en otro conjunto de datos distinto.

El Transfer Learning es una técnica de Machine Learning que permite utilizar un modelo desarrollado para una tarea específica como punto de partida para otra tarea distinta (aunque relacionada). Además de permitirnos obtener clasificadores de forma mucho más rápida aprovechando el conocimiento previo, el Transfer Learning hace posible el uso del Deep Learning con conjuntos de datos pequeños con los que sería imposible entrenar una red desde 0. El Transfer Learning es considerado por muchos investigadores como un paso más en dirección hacia la AGI.⁷

3.3.1. TRANSFER LEARNING CON IMÁGENES

En la práctica, cada vez es menos común el entrenamiento de redes convolucionales *from scratch*. Existen 2 principales motivos:

⁷Artificial General Intelligence: Aquella inteligencia artificial que puede realizar con éxito cualquier tarea intelectual de cualquier ser humano

- En determinados ámbitos, no siempre existen datasets con una gran cantidad de imágenes, suficiente para entrenar una red convolucional desde cero.
- Aún existiendo dicho dataset, el tiempo necesario para su completo entrenamiento puede ser de días, semanas o incluso meses dependiendo del equipo usado, la cantidad de datos y la complejidad de la arquitectura de la red.

Existen tres principales estrategias a la hora de realizar Transfer Learning:

- **Red convolucional como extractor de características:** Como se ha analizado anteriormente, una red convolucional puede ser vista como una herramienta para extraer características de las imágenes que posteriormente serán usadas por capas *fully connected* (o por cualquier otro tipo de clasificador) para realizar la clasificación. Conociendo esto, podemos utilizar la red convolucional entrenada para un conjunto de imágenes en otro conjunto de imágenes distinto, siendo el clasificador final el único que tendrá que ser reentrenado.
- **Fine-tuning de la red convolucional** Como se ha analizado anteriormente, las capas iniciales de las redes convolucionales se encargan de detectar características más generales y patrones simples, que van siendo más complicados a medida que avanzamos hacia capas posteriores. Por lo tanto, es común que estas primeras capas tengan siempre contenidos similares incluso en modelos entrenados con diferentes conjuntos de imágenes. Estas capas podrán ser reaprovechadas, con lo que únicamente tendremos que reentrenar las últimas capas y el clasificador final.
- **Modelos pre-entrenados:** Este tercer caso supone el reentrenamiento total de la red, sin embargo, partiendo de unos pesos que han sido previamente entrenados en otro conjunto de imágenes. De esta forma se consigue que el número de iteraciones necesarias hasta llegar al nivel de exactitud requerido sea menor.

Los criterios para decidir qué estrategia de Transfer Learning usar en cada caso dependen principalmente de las diferencias de contenido y tamaño entre

las imágenes de nuestro dataset y las del dataset original (con el que se entrenó el modelo que vamos a reutilizar)

Es común usar las siguientes *rules of thumb* como guía en función de 4 posibles escenarios:⁸

- **El nuevo dataset es pequeño pero similar al original:** Al tratarse de un dataset pequeño, modificar las capas convolucionales de nuestro modelo original puede dar lugar a **sobreajuste**. Por lo tanto, y puesto que las imágenes de ambos datasets son similares, la estrategia adecuada será utilizar la red convolucional como extracto de características y entrenar únicamente el clasificador final.
- **El nuevo dataset es grande y similar al original:** En este caso, como tenemos más imágenes podremos realizar fine-tunning de la red sin miedo a caer en sobreajuste.
- **El nuevo dataset es pequeño y muy diferente al original:** De nuevo, al tener un dataset pequeño, descartaremos entrenar la red convolucional. En este caso, lo que haremos es entrenar solo un clasificador. Además, al ser las imágenes distintas a las del dataset original, no podremos aprovechar las últimas capas de la red convolucional que serán eliminadas.
- **El nuevo dataset es grande y muy diferente al original:** En este caso entrenaremos la red convolucional al completo. Sin embargo, será de utilidad comenzar nuestro entrenamiento a partir de un modelo pre-entrenado.

⁸<http://cs231n.github.io/transfer-learning/>

3.4. Explicabilidad las redes convolucionales

Para la introducción de técnicas basadas en Deep Learning en diversos sectores profesionales (siendo la medicina uno de ellos), tan importante como la exactitud de las predicciones, es la existencia de técnicas que permitan **explicar el por qué de esas predicciones** o, al menos, dar un valor de **confianza** para cada una. Aunque las redes neuronales sean generalmente consideradas como **cajas negras**, en los últimos años han aparecido nuevas técnicas que permiten entender los factores que han llevado al clasificador a tomar una u otra decisión. Concretamente, durante este trabajo, se hará uso de una técnica para la interpretación de redes neuronales convolucionales conocida como **Mapas de Atención**, técnica basada en **Gradient-weighted Class Activation Mapping (Grad-CAM)** (Selvaraju et al. 2017). Gracias a estos mapas podemos conocer cuáles son las zonas de una imagen que más han influido en una predicción. Para ello, esta técnica usa los gradientes específicos de cada clase que fluyen hasta la última capa convolucional para producir un mapa de calor con las zonas de interés para la detección de esa clase (Figura 3.9).

Esta técnica no requiere modificar la arquitectura de la red ni volver a entrenarla. Además, permite obtener la localización aproximada de los objetos detectados en la imagen, aunque durante el entrenamiento no se haya utilizado ningún tipo de información de localización.

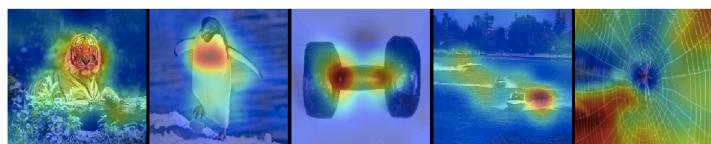


Figura 3.9: Mapas de atención generados por Grad-Cam para distintas clases de Imagenet.
Fuente: <https://github.com/raghakot/keras-vis>

3.5. Aplicaciones médicas del Machine Learning

¿Cómo sería un sistema sanitario en el que cada decisión relacionada con una enfermedad, en lugar de ser tomada por una sola persona, fuera tomada por un conjunto de los principales expertos del mundo de esa enfermedad? Esa es la pregunta que se hacen multitud de investigadores (Rajkomar et al. 2019). Estos concluyen que los tratamientos recetados de esta forma, y no los más conocidos por una única persona que los prescribe, serían los más efectivos. Además, se evitaría el **error humano**. Por desgracia, un sistema de este tipo sería inviable debido principalmente a la falta de expertos, que no darían abasto para diagnosticar a millones de pacientes cada día. Sin embargo, el Machine Learning nos promete un sistema similar a este pero realmente viable y escalable, con la capacidad de **aplicar todas las lecciones recogidas de la experiencia colectiva** en cada una de las decisiones, sin que esto genere una gran carga de trabajo para unos pocos expertos.

Hace ya 50 años se ponía de manifiesto la necesidad de “*aumentar, o incluso remplazar las funciones intelectuales de los médicos*” (Schwartz 1970). Además, la implementación de los **Historiales Clínicos Electrónicos** en diversidad de sistemas de salud, proporciona una ingente cantidad de datos que podrían ser de gran utilidad para la creación de modelos de Machine Learning de todo tipo.⁹

El uso de herramientas estadísticas en medicina no es ninguna novedad. Desde antes de la irrupción de las técnicas más novedosas de Machine Learning y Deep Learning, la estadística descriptiva tenía un papel fundamental, estando prácticamente siempre presente en los artículos de las revistas de medicina. Son necesarias técnicas estadísticas que nos permitieran estudiar la eficacia de los fármacos o los factores de riesgo de determinadas enfermedades.

La rama de la **epidemiología**, cuyos orígenes se sitúan hacia el siglo IV a.C., trata de recopilar y tratar los datos de los pacientes y sus patologías, para estudiar la frecuencia y distribución de los diversos fenómenos relacionados con la salud. La epidemiología trata de encontrar patrones en las enfermedades

⁹Aunque no hay que olvidar las limitaciones derivadas de la privacidad y la protección de datos

centrándose principalmente en tres aspectos: tiempo, lugar y persona. Gracias a ella somos capaces de definir los problemas de salud más importantes de una comunidad, además de sus factores de riesgo. Con esta información podremos desarrollar programas de prevención o control, e incluso predecir tendencias de una enfermedad.

Sin embargo, la revolución del Machine Learning y el Deep Learning de los últimos años se empieza a hacer notar, aunque de forma más lenta que en otros campos, en la medicina. Por primera vez, este tipo de técnicas salen del ámbito de la investigación y son utilizadas para el **diagnóstico**. Tradicionalmente los programas utilizados en diagnóstico eran **sistemas expertos**. Este tipo de programas simplemente se limitaban a pedir una serie de datos sobre el paciente, y obtenían conclusiones a partir de una serie de reglas que previamente habían tenido que ser definidas por especialistas. Sin embargo, con sistemas basados en Machine Learning, **estas reglas son automáticamente inferidas a partir de datos históricos** (Figura 3.10). Una de las principales características del Machine Learning, que le hace destacar sobre otros métodos tradicionales, es su capacidad de manejar enormes cantidades de predictores y encontrar complicados patrones en ellos.

Además, debido a la gran cantidad de información no estructurada existente (imágenes, señales, textos, etc.) en medicina, como era de esperar, el **Deep Learning** puede jugar un papel esencial, permitiendo que los datos *hablen por sí mismos*. Sin embargo, en todo momento tenemos que tener presente que nuestras evaluaciones pueden ser demasiado optimistas o que el sobreajuste puede hacer que nuestros modelos dejen de funcionar al ponerlos en producción. Tener una Inteligencia Artificial explicable, de la que no solo

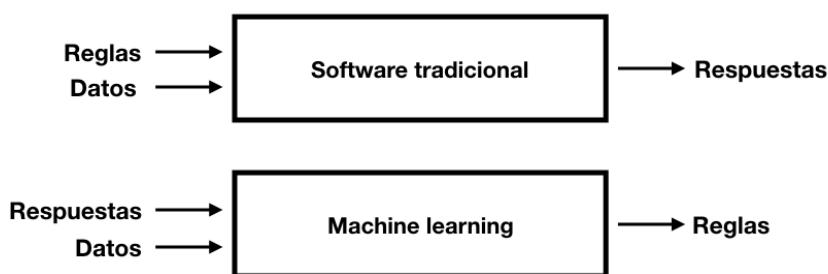


Figura 3.10: Diferencias entre Software tradicional y Machine learning. Elaboración propia.

obtengamos predicciones sino el por qué de las mismas, es algo que facilitará la entrada de estos algoritmos en el día a día de los médicos.

A continuación se detallarán los ámbitos dentro del campo de la medicina en la que el Aprendizaje Automático puede realizar importantes contribuciones.

3.5.1. PRONÓSTICO

El Machine Learning nos puede ayudar, mediante la búsqueda de patrones, en la predicción de la evolución de un enfermo. En varios servicios de salud existen ya implantados sistemas que, mediante Machine Learning, son capaces de identificar a los pacientes que están en riesgo de tener que ser transferidos a las unidades de cuidados intensivos. (Escobar et al. 2016). Además, diversos estudios sugieren que se pueden crear eficaces modelos de pronóstico médico a partir de la información en bruto de historiales (Rajkomar et al. 2018) e imágenes médicas (De Fauw et al. 2018).

La estandarización de los historiales médicos electrónicos sería de gran ayuda para la implantación de estos sistemas permitiendo, además, la agregación de datos. Formatos como el **Fast Healthcare Interoperability Resources (FHIR)** (Mandel et al. 2016) han nacido en los últimos años con este propósito.

3.5.2. DIAGNÓSTICO

Según concluye la Academia Nacional de Ciencias de EEUU, prácticamente todos los pacientes serán diagnosticados de forma errónea al menos una vez en su vida (Ball et al. 2015). Diversos estudios han encontrado problemas sistemáticos en los servicios de salud de todo el mundo. Hay evidencias de que, en los sistemas en los que los servicios de diagnóstico y tratamiento los realiza una misma organización obteniendo mayores ingresos la compañía mediante la prescripción de medicamentos y la solicitud de nuevas pruebas médicas, la tendencia a hacerlo aumenta considerablemente (Currie et al. 2014).

Los datos históricos pueden ser de gran ayuda para la identificación de posibles patologías durante las visitas clínicas, encontrando patrones complejos y ayudando a eliminar posibles errores y sesgos humanos. Los modelos podrían, incluso, sugerir nuevas pruebas a los médicos en base a los datos recogidos en tiempo real (Slack et al. 1966).

3.5.3. TRATAMIENTO

La aproximación más directa al problema del tratamiento mediante Machine Learning sería la creación de modelos entrenados con datos históricos que aprendieran los medicamentos recetados por los médicos en cada situación. Sin embargo, esta aproximación tiene un claro problema, el modelo aprendería los hábitos de prescripción de los médicos, que no tienen por qué ser los ideales. Por lo tanto, en este campo aún más, es de vital importancia generar datasets fiables y analizados en profundidad por expertos para entrenar los modelos (Rajkomar et al. 2019).

3.5.4. RETOS CLAVE

Uno de los principales retos en la creación de modelos de Aprendizaje Automático para medicina es la **falta de datos de calidad**. Este tipo de modelos, sobre todo los de Deep Learning, funcionan mejor cuanto mayor es la cantidad de datos de los que disponen para su entrenamiento. Sin embargo, en el campo de la medicina no existe tanta disponibilidad de los mismos como sí que existe en otros ámbitos. Una de las principales causas de esa escasez es la inviolable **privacidad** de los datos de los pacientes, que a menudo impide la creación de grandes datasets y únicamente permite crear conjuntos de datos lo suficientemente agregados como para que no pueda obtenerse datos de una persona en concreto. (Rajkomar et al. 2019)

Otro de los retos es el **sesgo** existente en los datos. Toda actividad humana está influenciada por un sesgo, ya sea consciente o inconsciente. La máxima **Entrada basura/Salida basura** de la analítica de datos está presente también en este campo. De nada servirá contar con potentes modelos capaces

de aprender complicados patrones, si luego esos patrones los encontrará en datos erróneos o sesgados.

La **interpretabilidad** de los modelos es también clave. Los médicos deben conocer el grado de veracidad y las limitaciones de estas técnicas para poder incorporarlas como una herramienta más. La sobreconfianza en estos sistemas puede conllevar una disminución de la alerta de los médicos que puede tener consecuencias letales. Que los modelos proporcionen, junto con sus predicciones, un grado de confiabilidad es un buen principio, pero no basta. De hecho, en ocasiones estos intervalos de fiabilidad pueden ser interpretados de forma incorrecta (Jiang et al. 2018). Es necesario crear modelos que sean capaces de explicar el por qué de sus predicciones. De hecho, esto era uno de los requisitos que indicó la Unión Europea en su **Guía ética para una Inteligencia Artificial fiable**.¹⁰ La necesidad de interpretabilidad de los resultados pude suponer un problema en técnicas de Deep Learning, que siempre han sido tachadas de ser **cajas negras**. Sin embargo, en los últimos años se han realizado diferentes estudios que demuestran que los modelos de Deep Learning pueden ser interpretables con las herramientas adecuadas. (Cruz-Roa et al. 2013), (Lipton 2016), (Zhang & Zhu 2018).

3.6. Correlación no implica causalidad

Aunque sea un mantra repetido hasta la saciedad en la literatura, esta advertencia merece un apartado propio en un trabajo de estas características, pues es algo a tener en cuenta y que implica tener mucha cautela al obtener conclusiones mediante este tipo de métodos. En muchas ocasiones creemos, de forma errónea, que existe una relación de causa y efecto entre dos variables que están correlacionadas, cuando esto no siempre es cierto.

La correlación entre dos variables puede ser debida a una tercera **variable oculta** que no tenemos por qué conocer o simplemente puede ser lo que conocemos como **correlación espúrea**, es decir, mera casualidad (que no causalidad).

¹⁰<https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

Sin embargo, la falacia **Cum hoc ergo propter hoc** (en latín, “Con esto, por tanto a causa de esto”) sigue siendo estando muy presente en los medios de comunicación y en las **pseudociencias**.

Si alguna vez el lector divisa a un sujeto disfrazado de pirata, no lo tome por loco. Ese sujeto podría ser un seguidor de **Bobby Henderson**, creador de la iglesia pastafari que, cansado de argumentos de los creacionistas basados en esta falacia, realizó un estudio (Figura 3.11) en el que demostraba una clara correlación entre la temperatura global y el descenso del número de piratas (un claro ejemplo de la existencia de una variable oculta, el tiempo). Es común, desde entonces, que los seguidores de Henderson se disfracen de piratas para recordarlo.

Otro ejemplo curioso es la singular correlación entre el número de ahogados en piscinas en Estados Unidos y el número de apariciones en películas de Nicholas Cage (Figura 3.12), en este caso una clara correlación espúrea. Si se torturan los datos durante el tiempo suficiente, éstos confesarán lo que deseemos.

Sin embargo, lejos de quedar en una mera anécdota como las anteriores, es extremadamente preocupante que existan familias en todo el mundo que estén decidiendo no vacunar a sus hijos debido a una aparente correlación, en un estudio de 2010, entre el número de casos de autismo y las vacunaciones.

Por lo tanto, es necesaria una gran cautela antes de obtener conclusiones de los sistemas de Machine Learning. Además, no estaría de más, aunque no serán objeto de análisis en este trabajo, tener presentes el Sesgo del Superviviente¹¹ y la Paradoja de Simpson¹²).

¹¹https://es.wikipedia.org/wiki/Sesgo_del_superviviente

¹²https://es.wikipedia.org/wiki/Paradoja_de_Simpson

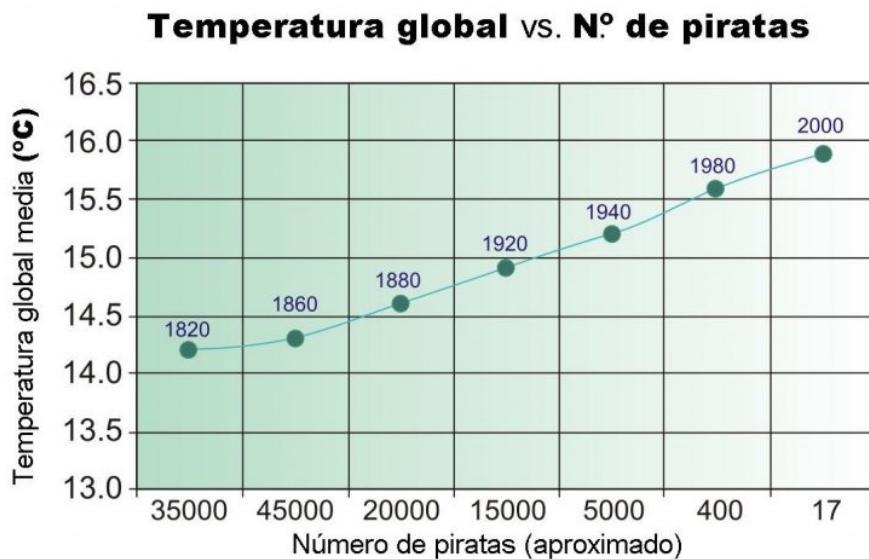


Figura 3.11: Correlación entre el aumento de la temperatura media global y el descenso del número de piratas. Fuente: <https://www.jotdown.es/2016/06/correlacion-no-implica-causalidad/>

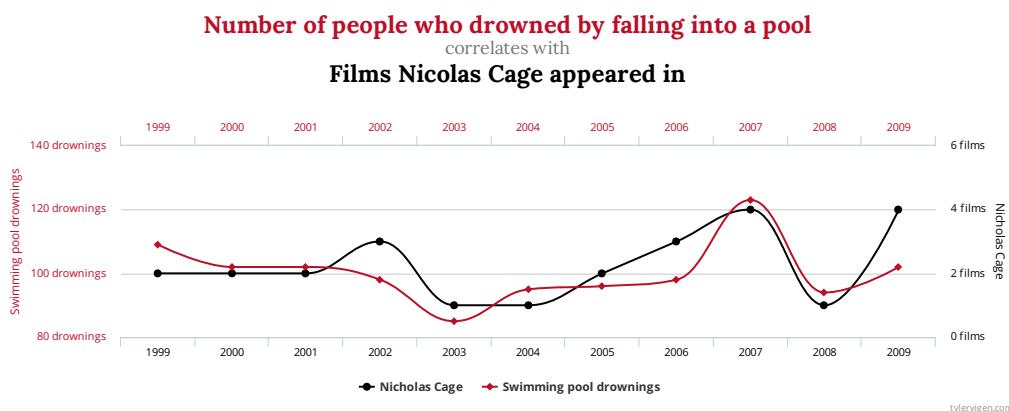


Figura 3.12: Correlación entre el número de ahogados en piscinas de Estados Unidos y el número de apariciones en películas de Nicholas Cage. Fuente: <http://www.tylervigen.com/spurious-correlations>

Capítulo 4

Estado del arte en detección de RD y DMAE

A lo largo de este capítulo se analizará el estado del arte de la detección de **Retinopatía Diabética** y **Degeneración Macular Asociada a la Edad** a partir de imágenes de fondo de ojo. En la actualidad, prácticamente la totalidad de los nuevos modelos publicados en este campo son modelos de **Deep Learning**. Sin embargo, se comenzará analizando los modelos basados en **Machine Learning** que precedieron a los actuales.

Al no existir en este campo un conjunto de datos de referencia sobre el que se evalúen todos los modelos, hay que tomar con cierta cautela las métricas de evaluación de los algoritmos que se detallarán a lo largo de este capítulo, puesto que cada modelo habrá sido entrenado y evaluado con datos distintos. De hecho, muchos de estos modelos han sido entrenados con conjuntos de apenas 100 o 200 imágenes, lo que hace muy probable que exista **sobreajuste** (*u overfitting*) y que el resultado de ponerlos en producción sea mucho peor del esperado.

4.1. Aproximaciones basadas en Machine Learning

Las modelos basados en Machine Learning para la detección de patologías en imágenes de fondo de ojo requieren una gran cantidad de **características**, escogidas y extraídas, de cada imagen, de forma manual por los investigadores. Para la obtención de las mismas es necesario **conocimiento experto** en la materia. Además, este tipo de características tienden a no generalizar bien, dando lugar a modelos, en muchas ocasiones, sobreajustados. Sin embargo, en los casos en los que nuestro conjunto de imágenes de entrenamiento es muy limitado, es común que los modelos basados en Machine Learning nos ofrezcan mejores resultados que los basados en Deep Learning. Además, la interpretabilidad de los modelos finales y sus predicciones es mayor en los modelos de Machine Learning que en los de Deep Learning.

4.1.1. DETECCIÓN DE RD MEDIANTE MACHINE LEARNING

Este tipo técnicas se basan en la búsqueda, en las imágenes, de cada una de las lesiones que caracterizan la Retinopatía Diabética. Las lesiones que caracterizan la RD, como se ha analizado anteriormente, son: **exudados**, **microaneurismas** y **hemorragias**. En el caso de la PDR, es posible encontrar también **neovascularización**. A partir de características obtenidas principalmente mediante técnicas de procesamiento digital de imagen y gracias al uso de **clasificadores basados en Machine Learning** es posible detectar la enfermedad, e incluso, estimar su gravedad.

Muchos de estos modelos comienzan por la obtención de imágenes binarias que representaran los **vasos sanguíneos** presentes en la imagen de la retina (Figura 4.1). La longitud, tamaño o posición de los mismos son de gran ayuda para el diagnóstico de la RD. Mediante la aplicación de una serie de técnicas al canal verde de las imágenes de fondo de ojo, es posible aislar estos vasos del resto de la imagen (Acharya et al. 2009). Otros modelos se basan también en la detección y seguimiento de las líneas centrales de los vasos sanguíneos (Tolias & Panas 1998) (Englmeier et al. 2004) (Vlachos & Dermatas 2010). También existen técnicas más avanzadas para ello basadas en el uso de **filtros adaptados** de dos dimensiones (Katz et al. 1989) (Hoover

et al. 1998) (Mookiah, U Rajendra Acharya, Martis, et al. 2013) (Gang et al. 2002). A partir de estos pre-procesamientos, existen sistemas capaces de detectar anchuras anormales en estos vasos sanguíneos (Hayashi et al. 2001), que suelen ser un indicio de la existencia de RD.

La presencia de **hemorragias** en las imágenes de fondo de ojo es mayor en los estadios más graves de la enfermedad. Su detección se realiza habitualmente junto con la detección de los vasos sanguíneos.

La presencia de **exudados** es el síntoma más característico de RD. Para la detección de éstos es común comenzar por la eliminación de los vasos sanguíneos y el disco óptico de las imágenes. Una vez eliminados estos elementos, es posible detectar los exudados mediante una secuencia de algoritmos de procesamiento de imagen (Acharya et al. 2009). Técnicas más avanzadas, basadas en **clasificadores estadísticos** basados en los niveles de brillo y uso de ventanas espaciales como estrategia de verificación han obtenido resultados cercanos al 100 % de exactitud en la detección de imágenes con exudados (sensibilidad) y 70 % de exactitud en la detección de imágenes de retinas sanas (especificidad) (Wang et al. 2000). Técnicas de detección de exudados basadas en **redes neuronales** (Hunter et al. 2000) o en el algoritmo **PCA**¹ (Li & Chutatape 2000) también han conseguido resultados similares a los comentados anteriormente. Esta última técnica, también permitía obtener la localización del disco óptico y la fóvea con unas exactitudes, respectivamente, de 99 % y 100 %.

A partir de la presencia de **microaneurismas** es también posible la detección de la RD, llegando a conseguirse una sensibilidad del 85 % y una especificidad del 90 % (Jelinek et al. 2006). La forma de detectarlos es similar a las anteriores y requiere eliminar el disco óptico y los vasos sanguíneos de la imagen antes de aplicar una serie de técnicas de procesamiento de imágenes. También es posible el uso de técnicas basadas en **morfología matemática** (Walter et al. 2007) (Hatanaka et al. 2008). Algunos de estos métodos realizan **transformaciones del espacio de color** de las imágenes a HSV (Hue, Saturation, Value), espacio donde es más fácil realizar el procesamiento. El uso de las **transformada Wavelet** también ha demostrado ser eficaz en

¹Principal Components Analysis

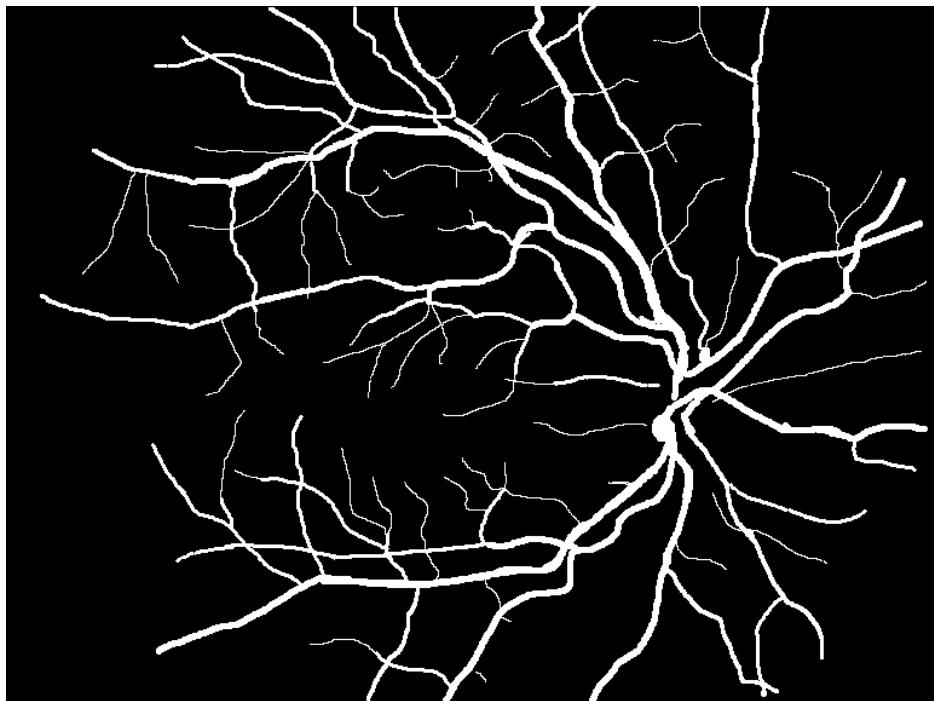


Figura 4.1: Imagen binaria de los vasos sanguíneos de la retina. Fuente: <http://www.aria-database.com/>

la detección de microaneurismas (Quellec et al. 2008).

De la misma forma que muchas de las técnicas explicadas hasta ahora basan su predicción en la detección de alguna de las lesiones típicas de la RD, también existen sistemas más complejos que son capaces de detectar, de forma simultánea, los tres tipos de lesiones y realizar predicciones en base a la presencia de cada tipo de lesión mediante clasificadores como **árboles de decisión** o **redes neuronales** (Ege et al. 2000), (Sinthanayothin et al. 2002), (Sinthanayothin et al. 2003), (Reza & Eswaran 2011). Técnicas de aprendizaje no supervisado como el **FCM**² también han demostrado ser eficaces en esta tarea (Osareh et al. 2002)

Todas las investigaciones analizadas anteriormente trataban de predecir una variable binaria, la presencia o no de retinopatía diabética. Sin embargo, otras investigaciones han tratado de detectar también el **tipo de RD** (Proliferativa o No Proliferativa). Estas técnicas han conseguido sensibilidad y especificidad de más del 95 % (Mookiah, U Rajendra Acharya, Martis, et al.

²Fuzzy c-means

2013) mediante el uso de **redes neuronales**. Otros trabajos han intentado distinguir, con éxito, hasta **5 grados de RD** (Acharya et al. 2008),(Acharya et al. 2009), (Acharya et al. 2012).

4.1.2. DETECCIÓN DE DMAE MEDIANTE MACHINE LEARNING

A pesar de ser la mayor causa de ceguera en países desarrollados (Wong et al. 2014), la detección de la **Degeneración Macular Asociada a la Edad (DMAE)** en imágenes de fondo de ojo no ha despertado tanto interés en la comunidad científica como la Retinopatía Diabética. Una extensa búsqueda en la bibliografía arroja un solo estudio (Pead et al. 2019) que analiza todos los métodos actuales de detección de DMAE basados en Machine Learning y Deep Learning. Ese estudio concluye que únicamente existen 14 publicaciones que abordan este tema.³

El principal signo de DMAE, como se ha analizado en capítulos anteriores, es la aparición de **drusas**, que pueden ser observadas en la imágenes de fondo de ojo como pequeños conjuntos de manchas blancas y amarillas. Por lo tanto, este tipo de modelos tratarán de buscar estas lesiones en las imágenes. En la Figura 4.2 vemos cuáles son las tareas principales en la detección de DMAE mediante Machine Learning.

Como se ha explicado en el anterior apartado, los algoritmos basados en Machine Learning suelen requerir de una fase de **pre-procesado y extracción de características**, dado que no son capaces de procesar la imagen *en bruto*. Puesto que las drusas son pequeñas regiones brillantes en las imágenes, algunos métodos han utilizado para su detección diversas características calculadas a partir de los **histogramas** de las imágenes. Además, es común aplicar una **ecualización del histograma** como paso previo a la extracción de características para obtener un mayor contraste en las imágenes (Hijazi et al. 2010), (Yalin Zheng et al. 2012), (Mookiah, U Rajendra Acharya, Koh,

³La búsqueda realizada filtra las publicaciones que no contengan una evaluación robusta del modelo, que no estén escritas en inglés o que no utilicen un método basado en las imágenes de fondo de ojo

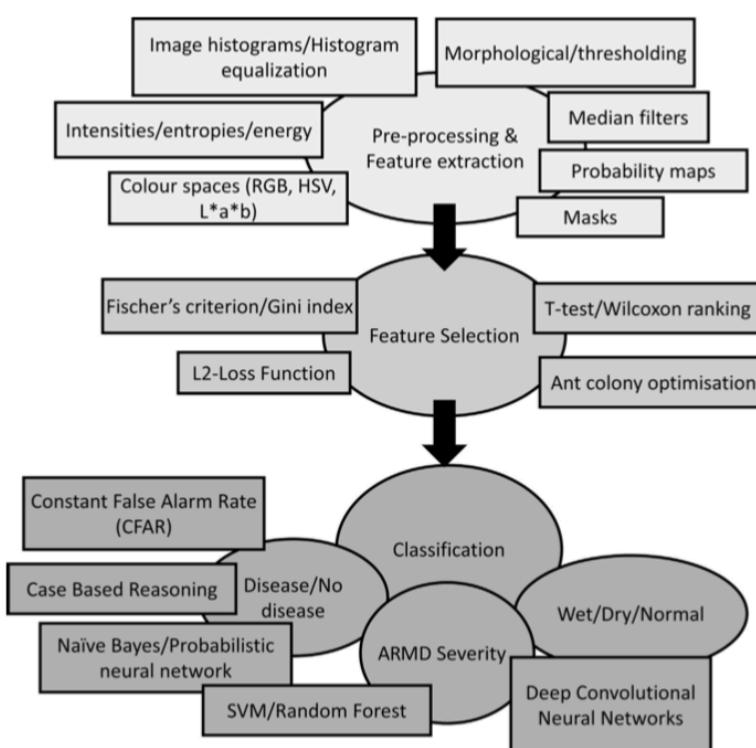


Figura 4.2: Resumen de las fases de la detección de DMAE mediante Machine Learning. Fuente: (Pead et al. 2019)

Chandran, et al. 2014), (Mookiah, U Rajendra Acharya, Koh, Chua, et al. 2014). Aplicar un **filtro de mediana** permite eliminar, antes de la extracción de características, el ruido de alta frecuencia de las imágenes (Kankanhalli et al. 2013) (Phan et al. 2016). El uso de técnicas de **morfología matemática** también ha demostrado ser eficaz para resaltar las regiones con drusas de la imagen (Burlina et al. 2011), (García-Floriano et al. 2017).

Tras el **pre-procesado y extracción de características** es común realizar una **selección de características** que nos permite eliminar características (o columnas o campos) que puedan resultar irrelevantes o puedan confundir a los algoritmos de Machine Learning. Para esto, se han utilizado algoritmos basados en la **correlación entre las distintas características** (García-Floriano et al. 2017) test paramétricos y no paramétricos como el **t-test** (Mookiah, U Rajendra Acharya, Koh, Chandran, et al. 2014), (Mookiah, U Rajendra Acharya, Koh, Chua, et al. 2014) o, incluso, **algoritmos genéticos** (Acharya et al. 2017)

En el último paso, la **clasificación**, podemos separar los algoritmos en dos grupos: los que simplemente tratan de diferenciar entre sano/enfermo y los que tratan también de detectar grados de afectación de la enfermedad. Dentro del primer grupo encontramos métodos como el **Razonamiento Basado en Casos** en el que, de forma similar al algoritmo **K-Nearest Neighbors**, simplemente se mide el grado de parecido entre el histograma de la imagen de fondo de ojo a predecir y cada uno de los histogramas de las imágenes del conjunto de datos de entrenamiento (Hijazi et al. 2010). También encontramos métodos basados en los algoritmos **Naive Bayes** o **SVM** (Zheng et al. 2011), (García-Floriano et al. 2017). Sin embargo, otras publicaciones han tratado de detectar también la gravedad de la enfermedad (estableciendo 5 posibles grados) con algoritmos como **Random Forest**, **K-Nearest Neighbors** o **SVM** (Kankanhalli et al. 2013), (Phan et al. 2016).

4.2. Aproximaciones basadas en Deep Learning

Las aproximaciones basadas en **Deep Learning**, haciendo uso de **Redes Neuronales Convolucionales**, representan actualmente el estado del arte

en multitud de tareas de análisis de imágenes médicas. Entre ellas se encuentra también, como no podía ser de otra forma, el análisis de imágenes de fondo de ojo. Una de las principales ventajas de este tipo de aproximaciones es que se elimina la necesidad de la extracción de características en base a conocimiento experto. De esta forma, **la red es directamente alimentada con las imágenes**, siendo tarea de ésta la extracción de características que permitan distinguir con eficacia las distintas clases de nuestro problema.

4.2.1. DETECCIÓN DE RD MEDIANTE DEEP LEARNING

Existen dos grupos principales de algoritmos de Deep Learning para la detección de Retinopatía Diabética: los que tratan de detectar y localizar en la imagen de fondo de ojo cada una de las lesiones típicas de la RD y los que, por el contrario, tratan de detectar directamente la presencia de la enfermedad sin enfocarse en detectar ni localizar las lesiones concretas. Es este segundo grupo el que nos interesa analizar, pues es la metodología que utilizaremos en la creación de nuestro modelo. Además, la gran ventaja de este tipo de modelos es que no necesita un dataset con anotaciones de la localización de las lesiones para ser entrenados. Simplemente necesitamos un conjunto de datos con etiquetas de **sano/enfermo**.

Prácticamente todos los modelos de este tipo han sido **Redes Neuronales Convolucionales**. La arquitectura **Inception** ha demostrado ser muy efectiva en la detección de la RD proliferativa. (Gulshan et al. 2016). En este caso, se trataba de una clasificación binaria donde sólo existían dos posibles salidas (enfermo/sano). Sin embargo, muchos otros investigadores han tratado de detectar diferentes niveles de gravedad (Colas et al. 2016), (Quellec et al. 2017), (Costa & Campilho 2017). Además, algunos de estos modelos son capaces de detectar también las lesiones concretas que aparecen en cada imagen (Colas et al. 2016), (Quellec et al. 2017)

Otros investigadores también han hecho uso, de forma satisfactoria, de la arquitectura **AlexNet** (Mansour 2018), o **ResNet** (Gargaya & Leng 2017). Además, este último modelo añadía al final de la red una capa convolucional adicional que permitía observar e interpretar el proceso de aprendizaje de la

red, solventando así el problema de interpretabilidad que a menudo tienen este tipo de modelos. La gran interpretabilidad, junto con una alta **accuracy** hacen que este modelo sea para muchos investigadores el **modelo de referencia** en detección de Retinopatía Diabética.

También se han realizado investigaciones con pasos previos de extracción de características usando técnicas como **Bag Of Visual Words** (Costa & Campilho 2017) o extracción del fondo de las imágenes mediante **Modelos Gaussianos Mixtos** (Mansour 2018)

Otras técnicas, como la de **Data Augmentation** consistente en crear nuevas imágenes a partir de transformaciones sobre las imágenes originales, han demostrado también ser eficaces (Pratt et al. 2016).

Además del Data Augmentation, otra técnica que también ha sido usada para solventar el problema de la falta de imágenes ha sido el **Transfer Learning**. Esta técnica ha sido aplicada utilizando, como base para nuestros modelos, otros modelos que habían sido previamente entrenados en otros datasets con todo tipo de imágenes (Maninis et al. 2016), (Li et al. 2017) o con datasets específicos de imágenes de fondo de ojo (Gondal et al. 2017). De ambas formas ha demostrado ser de utilidad, especialmente en los casos en los que el conjunto de imágenes de entrenamiento era demasiado reducido como para poder entrenar una arquitectura compleja desde cero.

4.2.2. DETECCIÓN DE DMAE MEDIANTE DEEP LEARNING

De la misma forma que con la Retinopatía Diabética, los métodos de detección de DMAE basados en Deep Learning han permitido saltar la etapa de extracción de características, delegándola en el propio clasificador. Prácticamente la totalidad de los modelos de Deep Learning de este tipo han hecho uso de **Redes Neuronales Convolucionales**. Estas redes han sido entrenadas desde 0 (Tan et al. 2018) o, en ocasiones se ha hecho uso de la técnica del **Transfer Learning**. Gracias a ésta, se han utilizado los pesos de redes entrenadas previamente en otros conjuntos de imágenes como punto de partida para el entrenamiento de las últimas capas de las redes convolucionales (Burlina et al. 2016). Además, como es común en Deep Learning, también

se han propuesto algunos modelos basados en la combinación de las predicciones de varios modelos distintos (Grassmann et al. 2018), técnica conocida como **ensemble**. Sin embargo, la cantidad de publicaciones que abordan la detección de DMAE mediante Deep Learning es aún muy limitada y es de esperar que muchas de las técnicas avanzadas que ya se están utilizando en la detección de Retinopatía Diabética sean aplicadas durante los próximos años a la detección del a Degeneración Macular Asociada a la Edad.

Capítulo 5

Diseño de Sistema de Detección de RD y DMAE

La gran cantidad de conjuntos de imágenes utilizados y el **extremo desbalanceo** de las clases han sido los dos factores que más han condicionado el diseño de los sistemas. Esto ha provocado que se hayan realizado **3 aproximaciones distintas al problema**, todas ellas basadas en Deep Learning, y únicamente obteniendo resultados útiles de 2 de ellas.

Durante este capítulo se analizarán estas aproximaciones. Este análisis no se limitará a una simple descripción del clasificador utilizado, sino que se detallarán los principales aspectos de cualquier proyecto de este tipo: los **datos** usados, su limpieza y procesado, el proceso de selección de **hiperparámetros** o incluso las características y limitaciones impuestas por los **recursos hardware y software** utilizados.

5.1. Exploración de los datos

Una de las principales contribuciones de esta investigación ha sido precisamente la **extensa cantidad de conjuntos distintos de imágenes** utilizados en la creación de los modelos. Para entrenar el modelo se han seleccionado imágenes de prácticamente todos los datasets utilizados por los sistemas del

capítulo 4. En total han sido utilizados **13 conjuntos de imágenes**, con un total de **39118 imágenes**.¹ Destaca el dataset **Kaggle** que contiene el 66 % del total de imágenes utilizadas. Este dataset proviene de una competición² realizada en 2015 que supuso importantes avances en la detección de Retinopatía Diabética a partir de imágenes de fondo de ojo.

Como se ha visto, la cantidad total de imágenes es muy elevada, siendo muy superior al tamaño medio de los datasets de los modelos analizados en el capítulo 4. Algunos de los modelos creados han utilizado grupos más reducidos de imágenes, seleccionadas aleatoriamente del conjunto de datos original. En la tabla 5.1 se muestra la cantidad de imágenes de cada tipo existentes en cada uno de los datasets utilizados.

Tabla 5.1: Cantidad de imágenes de cada tipo en cada uno de los conjuntos de imágenes utilizados

Dataset	SANA	RD	DMAE
GRAND-CHALLENGE	311	0	89
ARIA	61	59	23
DIARET DB0	20	110	0
E-OPTHA	268	195	0
HEI-MED	0	169	0
HRF	15	15	0
KAGGLE	25810	9316	0
MESSIDOR	540	660	0
ONHSD	0	99	0
ROC	0	50	0
DIAGNOS	23	0	22
STARE	37	89	47
FOM	533	457	101

Haber utilizado todos estos datasets ha supuesto una dificultad añadida al proceso pues se ha tenido que realizar un costoso trabajo previo de **selección**.

¹Sin embargo, como se verá durante este capítulo, algunos de los clasificadores que se han entrenado no han utilizado el conjunto completo de imágenes para el entrenamiento.

²<https://www.kaggle.com/c/diabetic-retinopathy-detection/>

ción, limpieza y preparación de los datos. Se han creado una serie de scripts que han recorrido cada una de las carpetas y han separado las imágenes de cada tipo.

Como se puede observar en los datos de la tabla 5.2, el gran problema del conjunto de imágenes utilizado, que ha condicionado en gran medida la forma de trabajar con él, es el gran **desbalanceo** existente entre las clases. La clase predominante, las imágenes de retinas sanas, contiene **más del 70 %** del total de imágenes. Por el contrario la clase minoritaria, las imágenes de retinas con DMAE, únicamente contiene 281 imágenes, **menos del 1 % del total**. Este gran desbalanceo nos obligará a aplicar diversas técnicas que permitan compensarlo como la **asignación de pesos distintos a las instancias de cada clase** en el cálculo de la función de coste o el **submuestreo de los datasets**.

Tabla 5.2: Cantidad de imágenes de cada tipo en el conjunto completo de datos utilizado

Clase	Total de imágenes	% del dataset completo
Todas	39118	100
Sanas	27618	70.60
RD	11219	28.68
DMAE	281	0.72

Otra dificultad derivada del uso de 13 datasets distintos es que nuestro clasificador tendrá que tratar imágenes con características muy distintas, como se observa en la tabla 5.3. Las condiciones en las que han sido tomadas, procesadas y almacenadas las imágenes varían en gran medida entre los distintos datasets. Sin embargo, si se pretende crear un clasificador robusto que sea capaz de trabajar en todo tipo de condiciones, utilizar esta elevada cantidad de conjuntos de imágenes será de gran ayuda.

Tabla 5.3: Características de las imágenes de cada uno de los conjuntos utilizados

Dataset	Origen	Tamaño	Formato
GRAND-CHALLENGE	(Anón s. f.)	Varios	JPG
ARIA	(Yalin Zheng et al. 2012) (Farnell et al. 2008)	576x768	TIFF
DIARET DB0	(Kauppi et al. 2006)	1500x1152	PNG
E-OPTHA	(Decencière et al. 2013)	Varios	JPG
HEI-MED	(Giancardo et al. 2012)	2196×1958	JPG
HRF	(Odstrcilik et al. 2013)	3504×2336	JPG
KAGGLE	(Cuadros & Bresnick 2009)	Varias	JPG
MESSIDOR	(Decencière et al. 2014)	2240×1488	TIFF
ONHSD	(Lowell et al. 2004)	760×570	BMP
ROC	(Niemeijer et al. 2009)	Varias	JPG
AMD DIAGNOS	Privada	Varias	JPG
STARE	(Hoover et al. 1998)	605x700	TIFF
FOM	Privada	Varias	JPG

Como puede verse en la tabla 5.3 algunas de las bases de datos utilizadas no son bases de datos públicas sino que han sido facilitadas por diversas instituciones al **Computer Vision and Behaviour Analysis Lab (CVBLab)**³ de la Universidad Politécnica de Valencia en el contexto del proyecto **Acríma**.⁴ Gracias a estas bases de datos privadas, se ha podido contar con un conjunto de imágenes de DMAE suficientemente grande como para aplicar técnicas de Deep Learning.

5.2. Recursos utilizados

Respecto al **software** utilizado, la librería Open Source **Keras**⁵ ha sido la elegida. Keras es una librería de Deep Learning de alto nivel en Python que

³<http://www.cvblab.webs.upv.es>

⁴http://www.cvblab.webs.upv.es/project/acrima_en/

⁵<https://keras.io/>

permite realizar, de forma rápida, todo tipo de redes neuronales. Además, es capaz de trabajar sobre varios frameworks de más bajo nivel: Theano,⁶ CNTK⁷ o **Tensorflow**⁸. Será éste último el framework sobre el que crearemos nuestros modelos con Keras (Figura 5.1). Otra importante característica de Keras a tener en cuenta es que permite la ejecución tanto en CPU como en GPU, sin necesidad de modificar para ello el código.

También se ha hecho uso de las Jupyter Notebooks⁹ (Figura 5.2), entornos de trabajo que permiten la creación de documentos que combinen fragmentos de código con texto, imágenes e incluso elementos interactivos. Las Jupyter Notebooks han ayudado a mostrar de forma simple y ordenada al usuario los resultados de las predicciones realizadas a partir de las imágenes de fondo de ojo proporcionadas.

En relación con el **hardware**, como todo proyecto de Deep Learning, la tarjeta gráfica utilizada durante el entrenamiento ha jugado un papel fundamental. En proyectos con grandes conjuntos de imágenes, como es el caso, no disponer de tarjeta gráfica (o disponer de una tarjeta sin la suficiente capacidad de procesamiento) puede hacer inviable el entrenamiento de los modelos. En este caso, la tarjeta gráfica utilizada ha sido la **NVIDIA TITAN Xp**. Esta tarjeta cuenta con una **arquitectura Pascal** con una frecuencia de reloj de 1481 MHz, y 12 GB de memoria. La potencia de cómputo de la NVIDIA TITAN Xp es de 12.15 TFLOPS. Además, parte del procesamiento en local se ha realizado en un **MacBook Air** con un procesador **Intel Core i7 de 2.2 GHz** y 8 GB de memoria RAM.

⁶<https://github.com/Theano/Theano>

⁷<https://github.com/Microsoft/cntk>

⁸<https://github.com/tensorflow/tensorflow>

⁹<https://jupyter.org/>

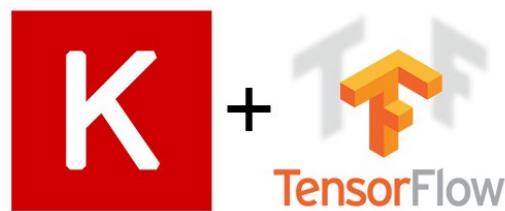


Figura 5.1: Logos de las dos librerías de Python utilizadas para la investigación: Keras y TensorFlow.

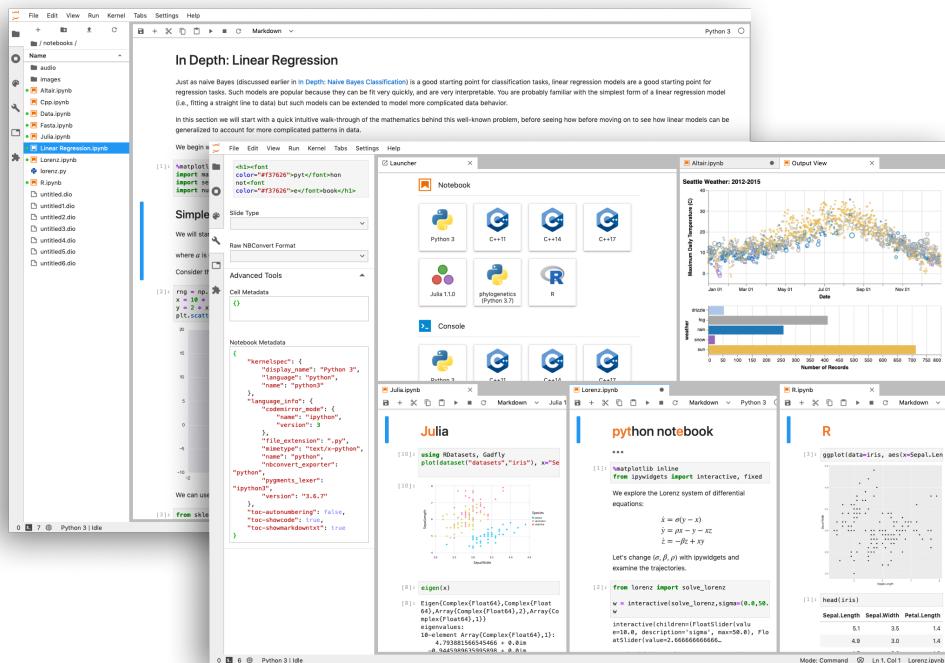


Figura 5.2: Ejemplos de Jupyter Notebooks. Fuente: <https://jupyter.org/>

5.3. Pre-procesado de las imágenes

Las siguientes transformaciones han sido aplicadas a las imágenes durante la fase de **pre-procesado**:

- **Reescalado del valor de los píxeles:** Dividiendo el valor de cada píxel entre 255 se obtienen únicamente valores entre 0 y 1, que son valores más comunes y, por lo tanto, más fáciles de manejar por los algoritmos de entrenamiento de las redes.
- **Reescalado de las imágenes:** Todas las imágenes han sido reescaladas a un tamaño de 224x224. La elección de este tamaño se debe a que es el **tamaño de las imágenes del dataset Imagenet**. Puesto que los pesos que usaremos como punto de partida de nuestra red vienen de una red entrenada con este dataset, mantener un mismo tamaño de imagen permitirá poder reusar mejor algunos de los filtros aprendidos.

La técnica conocida como **Data Augmentation** ha sido utilizada en todos los sistemas. Esta técnica permite añadir al conjunto de imágenes de entrenamiento, nuevas imágenes que provienen de la aplicación de diversas transformaciones a las imágenes del conjunto original. Esta técnica supondrá, como es obvio, un aumento del número total de imágenes del conjunto de entrenamiento y permitirá la obtención de clasificadores capaces de generalizar mejor ante nuevos casos. Dependiendo de las transformaciones utilizadas, el Data Augmentation permitirá la obtención de clasificadores más robustos frente al ruido, traslación/rotación de los objetos, a las variaciones del brillo, etc. En este caso, las transformaciones utilizadas han sido:

- Inversión del eje horizontal
- Aplicación de zoom aleatorio (hasta 1.25x)
- Desplazamiento aleatorio en el eje horizontal (hasta 10 %)
- Desplazamiento aleatorio en el eje vertical (hasta 10 %)
- Modificación aleatoria del brillo (entre -50 % y +50 %)

La librería utilizada, **Keras**, ha simplificado en gran medida este proceso. Las transformaciones elegidas y sus parámetros han sido escogidas de tal

forma que den lugar a imágenes *coherentes* como las que se podrían obtener con cualquier cámara de fondo de ojo.

5.4. Diseño del sistema 1: Gran clasificador

A continuación se presentarán las características de los sistemas de clasificación realizados. Estos sistemas tienen como finalidad la **detección de imágenes de retinas sanas, enfermedades de RD o enfermedades de DMAE**. Sin embargo, en ningún momento ha sido objetivo de este trabajo la detección de los diferentes niveles de gravedad de ambas patologías, principalmente debido a la falta de suficientes conjuntos de imágenes que proporcionen esta información para la fase de entrenamiento de los modelos. Los 3 sistemas presentados a continuación son totalmente independientes.

El primer sistema realizado (Figura 5.3) se trata de una **CNN basada en la arquitectura VGG16** (Simonyan & Zisserman 2014) que trata de distinguir, de una sola vez, entre los 3 tipos de imágenes (RD, DMAE, Sanas). A la salida de la última capa convolucional se han añadido 3 capas de tipo **fully connected** con 2048, 1024 y 512 neuronas. Entre ellas, para evitar el *overfitting* se han intercalado capas de tipo **Dropout**. Por último, la capa de salida cuenta con 3 neuronas (una por cada clase) y hace uso de la función de activación **softmax**. Para entrenar esta red se han utilizado las 39118 imágenes de todos los datasets descritos anteriormente.

El gran desbalanceo existente entre las clases, como se presentaba en la tabla 5.2, ha supuesto que este diseño no fuera capaz de detectar correctamente las 3 clases que componen nuestro problema y, por lo tanto, **ha sido desecharido**.

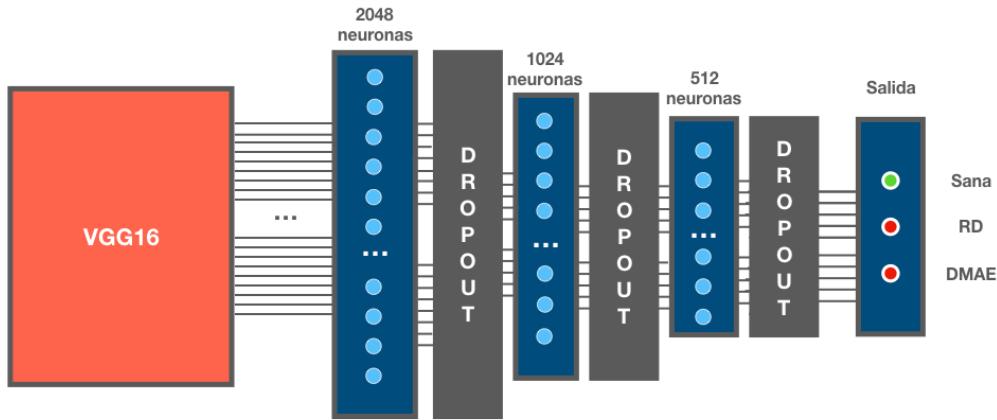


Figura 5.3: Arquitectura utilizada para el sistema 1. Elaboración propia

5.5. Diseño del sistema 2: Clasificador Multietapa

Los resultados del primer sistema ponen de manifiesto la necesidad de aplicar técnicas que traten el problema del desbalanceo. Por ello, el segundo sistema consta de **dos clasificadores binarios en cascada** (Figura 5.4):

- El primer clasificador diferencia entre **retinas sanas y retinas enfermas** (sin distinguir entre Retinopatía Diabética o Degeneración Macular).
- El segundo clasificador diferencia, de entre las imágenes de retinas detectadas como enfermas en el paso anterior, si el paciente está afectado de **RD o de DMAE**.

Gracias a este sistema basado en dos etapas se obtiene un desbalanceo entre clases, en cada una de las etapas, inferior al del conjunto original de imágenes. El primer clasificador **hace uso del conjunto completo de imágenes** para el entrenamiento pero, como se ha comentado, **únicamente es capaz de distinguir entre dos posibles casos, retinas sanas y retinas enfermas**. En la tabla 5.4 vemos la distribución de las imágenes de este primer clasificador en los conjuntos de entrenamiento, validación y test.

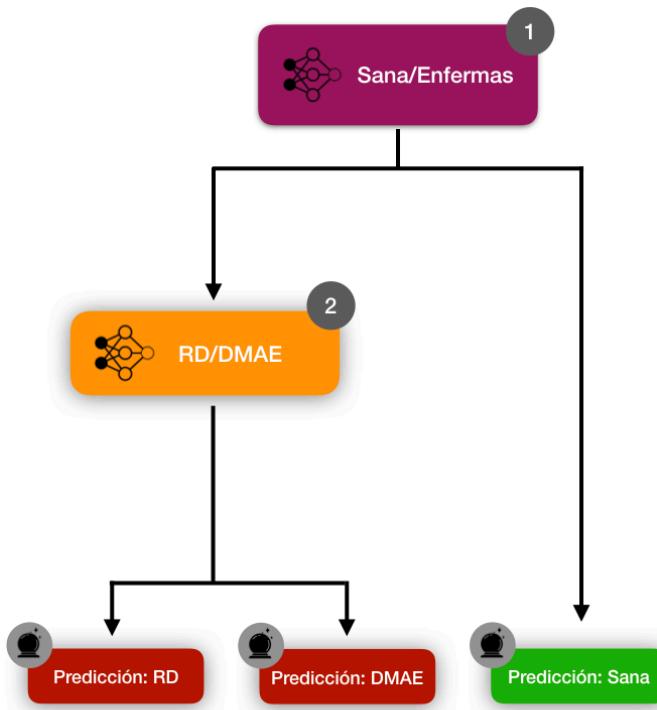


Figura 5.4: Arquitectura del sistema clasificador en dos etapas Elaboración propia

Tabla 5.4: Distribución de las imágenes del clasificador Sano/Enfermo del sistema 2.

Conjunto	Clase	Total	Total (%)
train	sana	17610	70.34
train	enferma	7425	29.66
valid	sana	4463	71.31
valid	enferma	1796	28.69
test	sana	5544	70.86
test	enferma	2280	29.14

Para tratar el **desbalanceo** existente en esta **primera etapa**, se ha hecho uso de una técnica basada en aplicar durante el entrenamiento a las instancias de la clase minoritaria (en este caso, la clase **enferma**) un peso que compense el desbalanceo en la función de coste.

Para la **segunda etapa** la aproximación ha sido distinta. Si se hubiera usado el conjunto de imágenes completo, el desbalanceo hubiera sido demasiado grande, imposible de abordar incluso por la técnica utilizada anteriormente. En este caso, se ha hecho uso de la técnica conocida como **subsampling o submuestreo**. Para evitar tener una cantidad de imágenes de RD muy superior a la de DMAE se han seleccionado, de forma aleatoria, un conjunto de imágenes de RD que serán las utilizadas para el entrenamiento. De esta forma, se entrenará el clasificador con la misma cantidad de imágenes de DR que de DMAE.

Esta arquitectura basada en dos etapas nos ha permitido usar la totalidad de las imágenes para el entrenamiento sin necesidad de entrenar modelos con datasets extremadamente desbalanceados (como era el caso del sistema inicial).

En la Figura 5.5 podemos ver la arquitectura utilizada en los clasificadores de ambos subsistemas. Como se puede comprobar, es prácticamente igual a la del sistema 1, pero en este caso se elimina una de las capas **fully connected**. El clasificador sano/enfermo únicamente ha hecho uso de la arquitectura **VGG16**, mientras que el clasificador RD/DMAE ha hecho uso de las 3 arquitecturas de la imagen: **VGG16** (Simonyan & Zisserman 2014), **Resnet50** (He et al. 2016) e **InceptionV3** (Szegedy et al. 2016). Puesto que ahora la salida de la red es binaria, se ha cambiado la función de activación softmax utilizada anteriormente en la última capa por la **función sigmoide**.

Ambos clasificadores han utilizado la técnica del **Transfer Learning**. Partiendo de los pesos originales procedentes de **Imagenet**, se han realizado diversos entrenamientos, cambiando el número de parámetros de la red *congelados*:

- Entrenamiento únicamente de las capas fully-connected
- Entrenamiento de las capas fully-connected y el bloque convolucional número 5: (Últimas 4 capas de la red convolucional)
- Entrenamiento de las capas fully-connected y los bloques convolucionales 4 y 5: (Últimas 8 capas de la red convolucional)
- Entrenamiento de las capas fully-connected y los bloques convolucio-

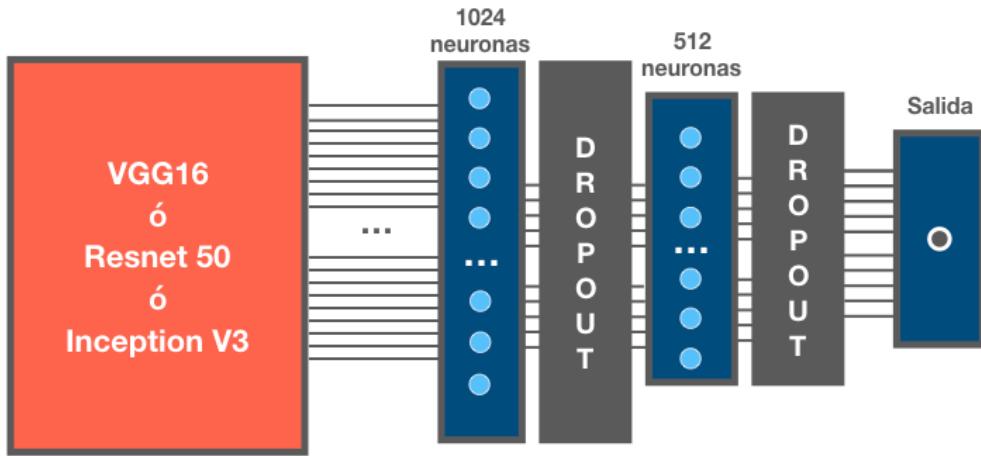


Figura 5.5: Arquitectura utilizada para los clasificadores de ambos subsistemas del segundo diseño.

- nales 3, 4 y 5: (Últimas 12 capas de la red convolucional)
- Entrenamiento de las capas fully-connected y los bloques convolucionales 2, 3, 4 y 5: (Últimas 15 capas de la red convolucional)
- Entrenamiento de la red completa

Como se detallará en el siguiente capítulo, la ejecución de varios entrenamientos alterando hiperparámetros como el *learning rate* o el *batch size* ha permitido obtener los valores óptimos para éstos.

5.6. Diseño del sistema 3: Ensemble de Clasificadores

El tercer sistema diseñado permite detectar los 3 posibles casos (RD, DMAE, y Sana) en una sola etapa a partir de la combinación de las predicciones de 3 clasificadores entrenados con diferentes subconjuntos de imágenes (Figura 5.6). De esta forma, al igual que en el caso anterior conseguimos entrenar modelos con la misma cantidad de imágenes en cada clase.

De la misma forma que el sistema anterior, este sistema también ha aplicado **Transfer Learning** descongelando progresivamente bloques de capas hasta llegar a obtener la mejor evaluación posible con el dataset de validación.

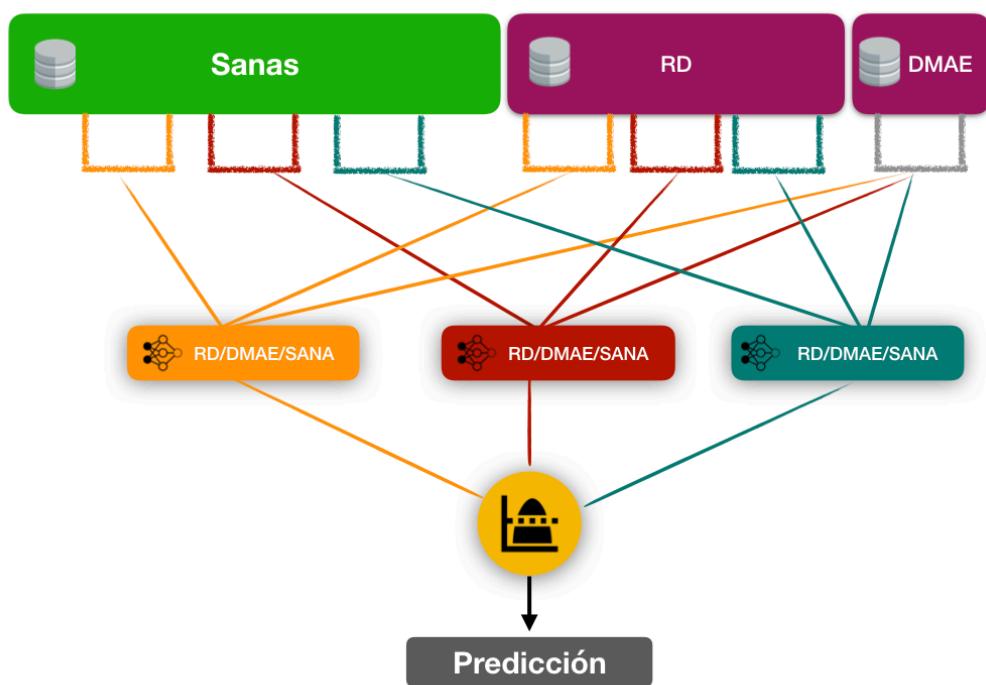


Figura 5.6: Arquitectura del sistema de ensemble de clasificadores simples. Los bloques superiores representan los conjuntos de imágenes utilizados para el entrenamiento de los mismos
Elaboración propia

La arquitectura utilizada ha sido la misma que la de la Figura 5.5 (aunque, en este caso, con 3 neuronas en la última capa, una por cada posible clase de salida), entrenándose con las mismas 3 arquitecturas: VGG16, Resnet50 e InceptionV3.

5.7. Diseño del Sistema de Predicción e Interpretación

Para hacer accesible el uso de los sistemas de predicción se ha diseñado un pequeño software que abstrae la complejidad de todo el proceso, permitiendo a los especialistas hacer uso de estos modelos simplemente seleccionando las imágenes del paciente que se desea analizar. Para ello, se ha hecho uso de las Jupyter Notebook explicadas anteriormente.

Mediante este programa, el usuario puede elegir el sistema que quiere utilizar en cada caso: el Clasificador Multietapa, o el Ensemble de Clasificadores. Una vez seleccionada la imagen de fondo de ojo a analizar, el programa devolverá sus predicciones, el grado de confianza y un mapa de calor procedente de la aplicación del algoritmo **Grad-Cam**.

En el capítulo 6, se analizan los resultados de diversas ejecuciones del programa.

Capítulo 6

Análisis de los resultados obtenidos

Durante este capítulo se analiza el funcionamiento de los sistemas descritos en el capítulo anterior. En todos los casos, el dataset original ha sido dividido en 3 subgrupos, obteniendo así:

- **Dataset de entrenamiento**
- **Dataset de validación**
- **Dataset de test**

Los resultados que se presentan durante este capítulo son los correspondientes a los datasets de validación y, en algunos casos, los de test, puesto que la evaluación del clasificador con el conjunto de datos de entrenamiento no es relevante para conocer la capacidad de **generalización** de los sistemas. El dataset de validación ha permitido evaluar el rendimiento de diferentes **arquitecturas e hiperparámetros** mientras que el dataset de test ha permitido dar, al final del proceso, una evaluación del sistema final con datos nuevos y comprobar el funcionamiento del **Sistema de Predicción e Interpretación**. Las métricas proporcionadas serán siempre las del **mejor epoch**, es decir, el que ha obtenido el mínimo valor de pérdidas con el dataset de validación.

Antes de pasar al análisis de los sistemas, es importante conocer cómo se obtienen las métricas utilizadas, como se presenta durante la siguiente sección.

6.1. Evaluación de sistemas de Machine Learning

Un paso tan importante como el modelado en un proyecto de análisis de datos es la evaluación de los resultados. Es de gran importancia establecer medidas que nos permitan saber cómo se está comportando nuestro modelo. En la literatura existe una gran cantidad de métricas, aunque en este caso nos centraremos en algunas de las más comunes en problemas de este tipo.

El problema analizado en este trabajo es un problema de **clasificación**, es decir la variable objetivo (la que predecimos) solo puede tomar un conjunto de valores discretos. Además, lo que inicialmente era un problema con tres posibles clases (RD/DMAE/Sano) también puede descomponerse en **dos problemas de clasificación binaria** (RD/Sano) (DMAE/Sano). Se trata de predecir una clase con sólo dos posibles valores. Cuando en un problema de este tipo comparamos la predicción realizada por un modelo con el *ground truth* (es decir, la clase que realmente correspondería a esa instancia), pueden darse 4 posibles casos:

- **Verdadero Positivo (o True Positive, TP)**: El sistema predice que el paciente **SÍ** tiene la enfermedad y acierta.
- **Verdadero Negativo (o True Negative, TN)**: El sistema predice que el paciente **NO** tiene la enfermedad y acierta.
- **Falso Negativo (o False Negative, FN)**: El sistema predice, erróneamente, que el paciente **NO** tiene la enfermedad cuando en realidad sí que la tiene.
- **Falso Positivo (o False Positive, FP)**: El sistema predice, erróneamente, que el paciente **SÍ** tiene la enfermedad cuando en realidad no la tiene.

A partir de la cantidad de predicciones de cada uno de estos posibles 4 tipos se pueden definir una serie de medidas muy comunes en problemas de este tipo.

La métrica más común, conocida como **Accuracy** o **Exactitud**, mide el porcentaje de aciertos del sistema (ecuación 6.1). Esta métrica carece de utilidad cuando tenemos conjuntos de datos desbalanceados.

$$\frac{TP + TN}{TP + TN + FN + FP} \quad (6.1)$$

La **Sensibilidad** mide la proporción de los pacientes que **Sí** tienen la enfermedad que nuestro clasificador ha sido capaz de detectar (ecuación 6.2)

$$\frac{TP}{TP + FN} \quad (6.2)$$

La **Especificidad**, en cambio, mide proporción de los pacientes que **No** tienen la enfermedad que nuestro clasificador ha sido capaz de detectar (ecuación 6.3)

$$\frac{TN}{TN + FP} \quad (6.3)$$

En función del campo de aplicación de los modelos, unas métricas toman más importancia que otras. Incluso es común tener **umbrales de actuación** en nuestros modelos que nos permitan elegir el punto de equilibrio deseado entre sensibilidad y especificidad. Un modelo que trata de predecir la presencia de una enfermedad siempre tratará de enfocarse más en obtener una buena **sensibilidad** antes de centrarse en la **especificidad**. El coste de predecir erróneamente que un paciente tiene una enfermedad, es menor al de no haber detectado la enfermedad en un paciente que sí que la tenía.

6.2. Evaluación del Sistema 1: Gran Clasificador

Como ya se anunciaba en el capítulo 5, debido al gran desbalanceo existente entre las clases, este sistema **no nos ha permitido distinguir correctamente entre las 3 clases**. Tras evaluarse varios valores distintos para el *learning rate* y *batch size* e incluso añadirse pesos a las clases que compensa-

ran el desbalanceo existente, el proceso de descenso de gradiente ha quedado constantemente *atrapado* en mínimos locales en los que el clasificador predice la misma clase para todas las instancias. Concluimos, por lo tanto, que **un solo clasificador no tiene capacidad suficiente para obtener patrones de un conjunto de datos tan desbalanceado** y habrá que buscar soluciones alternativas.

6.3. Evaluación del Sistema 2: Clasificador Multietapa

El segundo sistema consta de 2 etapas: la clasificación **Sano/Enfermo** y la clasificación **RD/DMAE**. Ambas etapas son analizadas a continuación.

6.3.1. ETAPA 1: CLASIFICADOR SANO/ENFERMO

Las 39118 imágenes de las que se componía nuestro conjunto de imágenes inicial han sido divididas en 2 grupos: **Retinas Sanas y Retinas Enfermas**.

Puesto que ha sido utilizada la técnica de **Transfer Learning**, se han realizado varias ejecuciones descongelando, progresivamente cada uno de los bloques convolucionales de la arquitectura utilizada, **VGG16**. Esta arquitectura posee 5 bloques convolucionales con **capas de pooling** en cada uno de estos bloques.

Inicialmente se han realizado varios entrenamientos del bloque *fully connected* para evaluar los posibles *batch size* y *learning rate*. La tabla 6.1 contiene los resultados obtenidos para distintos *batch sizes*. Las evaluaciones de la tabla son las correspondientes al dataset de validación para el mejor epoch.¹ El *learning rate* utilizado ha sido de 0.0001.

¹el que tiene menores pérdidas con el dataset de validación

Tabla 6.1: Resultados del entrenamiento para distintos batch size. Modelos evaluados con el dataset de validación

batch size	accuracy	loss
16	0.7062	0.5973
32	0.6782	0.6151
64	0.7072	0.6073

A partir de la tabla 6.1 se puede intuir que el tamaño del *batch size* no juega un papel de gran importancia en el proceso de entrenamiento por lo que se ha decidido usar, a partir de este momento un tamaño de **64**. Este tamaño nos permite entrenar la red de forma más rápida y nos asegura que en cada *batch* exista suficiente cantidad de imágenes de las dos clases. Usar tamaños superiores habría ocasionado problemas de memoria en la GPU utilizada.

De la misma forma, como se aprecia en la tabla 6.2, también se han realizado varios entrenamientos del clasificador que nos han permitido comprobar cuál es el *learning rate* adecuado para nuestro problema. Para ello se ha utilizado un *batch size de 64*.

Tabla 6.2: Resultados del entrenamiento para distintos learning rate. Modelos evaluados con el dataset de validación

learning rate	accuracy	loss
0.001	0.287	11.4
0.0001	0.6782	0.6151
0.00005	0.7199	0.6037

Como se ha podido comprobar, utilizar un *learning rate* demasiado alto provoca que el descenso de gradiente quede *atrapado* en mínimos locales o no sea capaz de converger, dando lugar a unos valores de pérdidas demasiado altos. El **learning rate de 0.0005** es el que nos da mejores resultados y por lo tanto ha sido utilizado en los posteriores entrenamientos. Este valor tendrá que disminuirse ligeramente cuando entrenemos varias capas convolucionales a la vez para asegurarnos un descenso de gradiente lento que pueda converger

en un mínimo absoluto.

Una vez decididos los hiperparámetros se ha comenzado a *descongelar* los diversos bloques de las capas convolucionales de la red, obteniendo los resultados de la tabla 6.3

Tabla 6.3: Resultados del entrenamiento para distintos bloques convolucionales entrenados. Modelos evaluados con el dataset de validación

train blocks	LR	accuracy	loss	sensitivity	specifity
FC	5e-5	0.7149	0.6549	0.1713	0.9338
Bloque 5	5e-5	0.7532	0.5294	0.3805	0.9012
Bloques 4,5	5e-6	0.7126	0.6854	0	1
Bloques 3,4,5	5e-6	0.7880	0.4748	0.4793	0.9131
Bloques 2,3,4,5	5e-6	0.7961	0.4704	0.5867	0.8624
Todos	5e-6	0.8015	0.4604	0.4680	0.9364

Como se puede ver en la tabla 6.3, **los mejores resultados se han obtenido al entrenar todos los bloques convolucionales**, obteniendo una *accuracy* de 80.15 %. Sin embargo, será la versión de la red en la que se han entrenado los bloques 2, 3, 4 y 5 la que se usará en el **Sistema de Predicción e Interpretación** para la segunda etapa del Clasificador Multietapa. Esta versión tiene una sensibilidad notablemente superior, con unas pérdidas similares a las de la red que ofrece las mínimas pérdidas. Cabe destacar que, a partir del entrenamiento del bloque 4, hemos tenido que disminuir el *learning rate* para evitar caer en mínimos locales y asegurar la convergencia.

Debido a la gran carga computacional que ha supuesto entrenar este modelo (cada entrenamiento ha durado una media de 72 horas), únicamente se ha evaluado la arquitectura **VGG16**.

La Figura 6.1 contiene la progresión de las pérdidas durante el entrenamiento del clasificador final elegido para esta etapa. **Las mínimas pérdidas se obtienen alrededor del epoch 45**. A partir de ese momento, la red empieza a sufrir de *overfitting* y las pérdidas con el dataset de validación aumentarán mientras que las del dataset de entrenamiento continuarán descendiendo. Esto es un claro indicador de que la red está comenzando a **memorizar** el

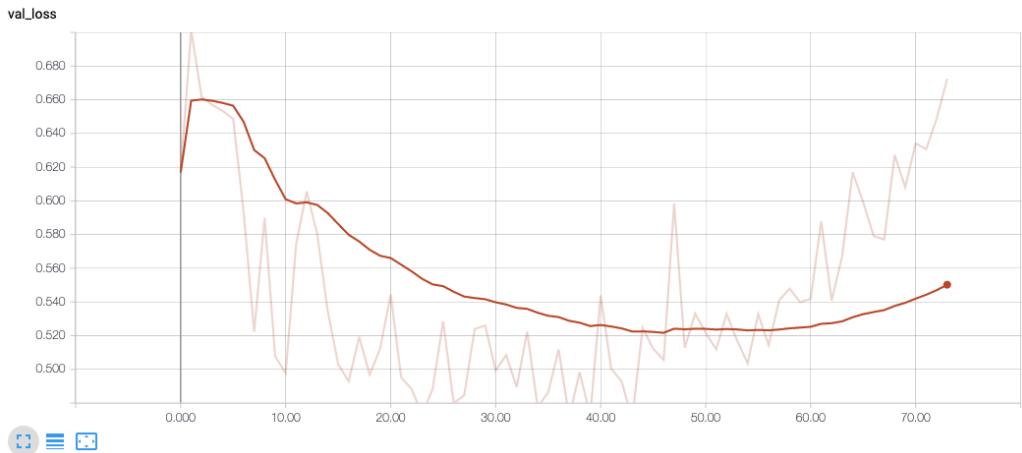


Figura 6.1: Pérdidas para el dataset de validación del entrenamiento de los bloques 2,3,4 y 5 (y FC) del clasificador Sano/Enfermo. Se ha aplicado un filtro de suavizado.

dataset de entrenamiento en vez de detectar y aprender patrones. Por lo tanto, nos quedaremos con el estado de la red en ese epoch 45.

En las Figuras 6.2, 6.3 y 6.4 se pueden ver ejemplos de la respuesta proporcionada por esta etapa del sistema.

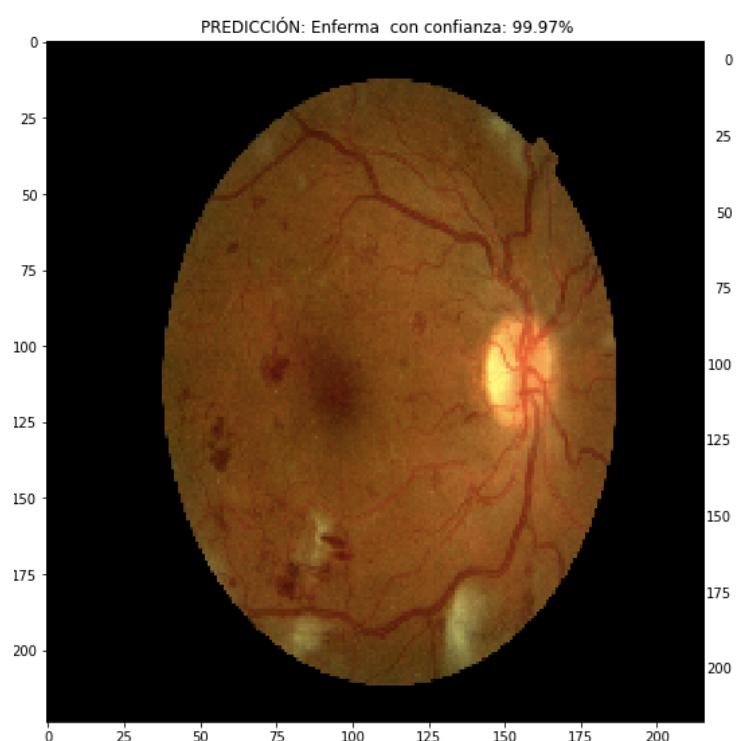


Figura 6.2: Salida de la primera etapa del Sistema Multietapa para una imagen de una retina enferma de RD

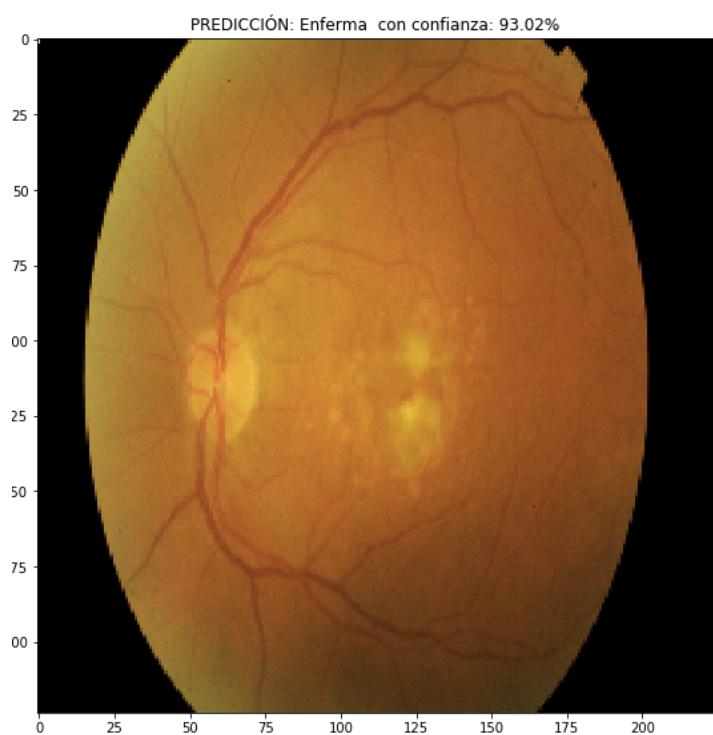


Figura 6.3: Salida de la primera etapa del Sistema Multietapa para una imagen de una retina enferma de DMAE

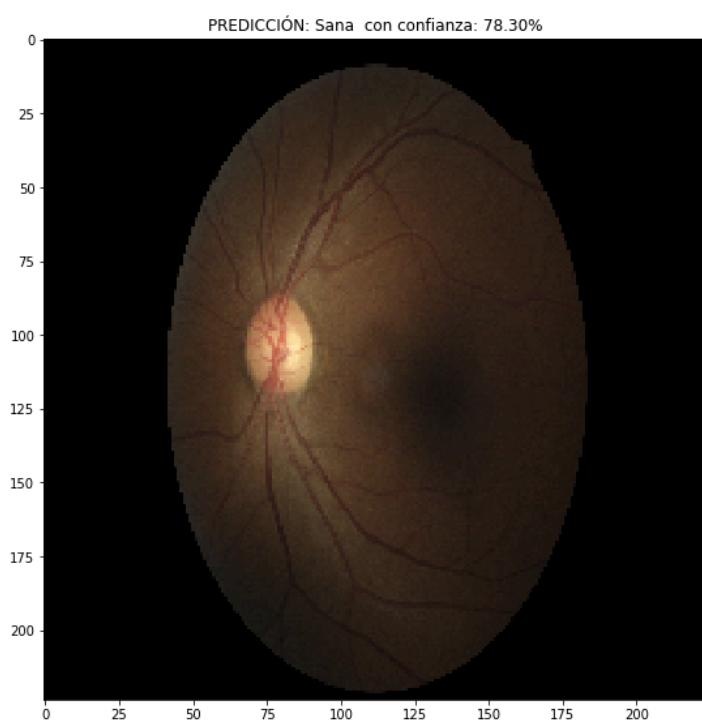


Figura 6.4: Salida de la primera etapa del Sistema Multietapa para una imagen de una retina sana

6.3.2. ETAPA 2: CLASIFICADOR RD/DMAE

La tabla 6.4 muestra los resultados del entrenamiento de la arquitectura **VGG16** para la **segunda etapa** del **Sistema Multietapa**. La función de este clasificador es diferenciar, de entre las imágenes de retinas detectadas como enfermas en la etapa 1, cuáles sufren Retinopatía Diabética y cuáles Degeneración Macular Asociada a la Edad.

Tabla 6.4: Resultados del entrenamiento de la segunda etapa del Sistema Multietapa. Modelos evaluados con el dataset de validación

train blocks	LR	accuracy	loss
FC	1e-5	0.9231	0.1910
Bloque 5	1e-5	0.9359	0.1483
Bloques 4,5	1e-5	0.8958	0.2093
Bloques 3,4,5	1e-5	0.9615	0.1443
Bloques 2,3,4,5	1e-5	0.9103	0.1773
Todos	1e-5	0.9487	0.1691

Además, como muestra la tabla 6.5, también se han entrenado otras arquitecturas. En este caso, en vez de realizarse *fine-tuning*, se han entrenado todos los bloques convolucionales de las mismas.

Tabla 6.5: Resultados del entrenamiento de la segunda etapa del Sistema Multietapa. Modelos evaluados con el dataset de validación

Arquitectura	LR	accuracy	loss
InceptionV3	1e-5	0.9167	0.263
Resnet50	1e-5	0.9744	0.0816

Los resultados en esta segunda etapa son mucho más satisfactorios que los de la primera, obteniéndose la máxima **accuracy** (97.4 %) con la arquitectura Resnet50, que será la utilizada en el **Sistema de Predicción e Interpretación**. Las Figuras 6.5 y 6.6 son dos ejemplos de la respuesta proporcionada por esta etapa del sistema.

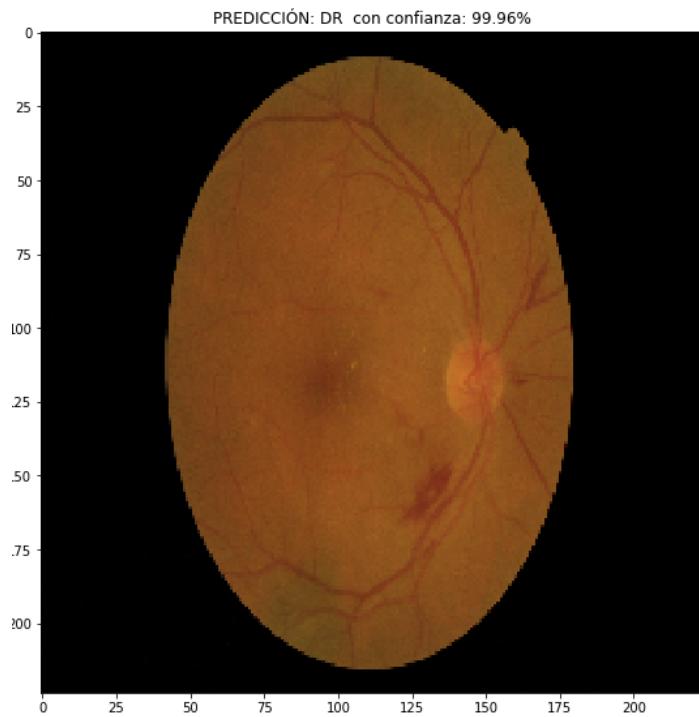


Figura 6.5: Salida de la segunda etapa del Sistema Multietapa para una imagen de una retina enferma de RD

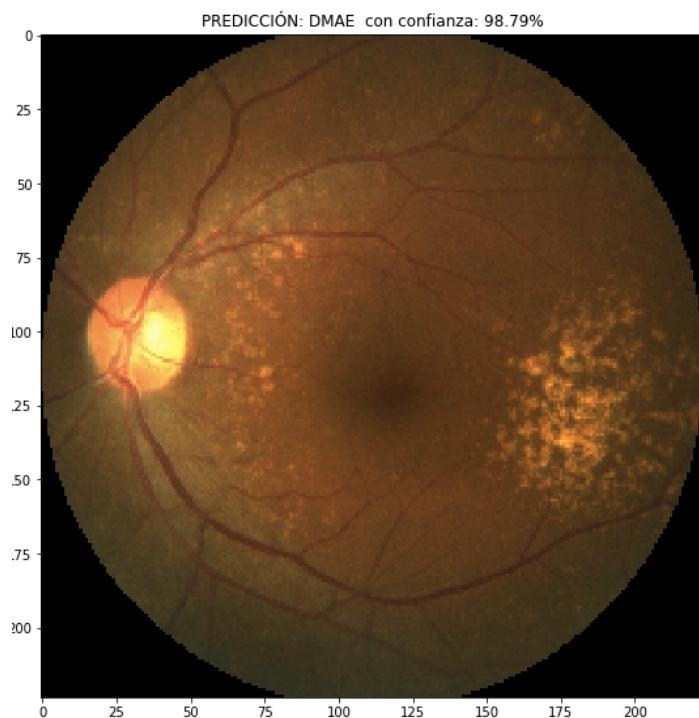


Figura 6.6: Salida de la segunda etapa del Sistema Multietapa para una imagen de una retina enferma de DMAE

6.4. Evaluación del Sistema 3: Ensemble de Clasificadores

Para este sistema se han evaluado 3 arquitecturas distintas: **VGG16**, **ResNet** e **InceptionV3**. Como se ha explicado anteriormente, cada una de estas tres arquitecturas ha sido entrenada con un subconjunto distinto de los datos. De esta forma se han obtenido clasificadores no correlados que, al ser combinados, han permitido obtener un rendimiento superior al de cada uno de ellos de forma individual.

Para la arquitectura VGG16, al utilizarse la técnica del **Transfer Learning**, se han evaluado diferentes versiones, congelando cada vez distinto número de capas como muestra la tabla 6.6.

Tabla 6.6: Resultados del entrenamiento del sistema 3. Modelos evaluados con el dataset de validación

train blocks	LR	accuracy	loss
FC	5e-5	0.6695	0.6676
Bloque 5	5e-5	0.7458	0.6068
Bloques 4,5	5e-6	0.7119	0.5878
Bloques 3,4,5	5e-6	0.7119	0.5776
Bloques 2,3,4,5	5e-6	0.7797	0.5456
Todos	5e-6	0.7458	0.5973

En este caso, el mejor resultado lo obtenemos dejando congelado el primer bloque convolucional y entrenando el resto de bloques.

En la tabla 6.7 se muestran los resultados del entrenamiento para otras arquitecturas.

Tabla 6.7: Resultados del entrenamiento con diferentes arquitecturas del sistema 3. Modelos evaluados con el dataset de validación

Arquitectura	LR	accuracy	loss
InceptionV3	1e-5	0.6076	0.7293
Resnet50	1e-5	0.6383	0.7260

El **ensemble** final usado en el **Sistema de Predicción e Interpretación** ha combinado las predicciones de las dos redes de la tabla 6.7 y la red que ha entrenado los bloques 2, 3, 4 y 5 de la tabla 6.6.

En las Figuras 6.7, 6.8 y 6.9 se pueden ver ejemplos de la respuesta proporcionada por este tercer sistema. Como se puede ver en ellas, existen ocasiones en que alguno de los 3 clasificadores del ensemble proporciona una respuesta equivocada. Sin embargo, el resultado final proporcionado por el ensemble devuelve la respuesta correcta. Es en estos casos donde se puede comprobar la robustez que proporciona haber usado un sistema basado en la combinación de varios modelos distintos.

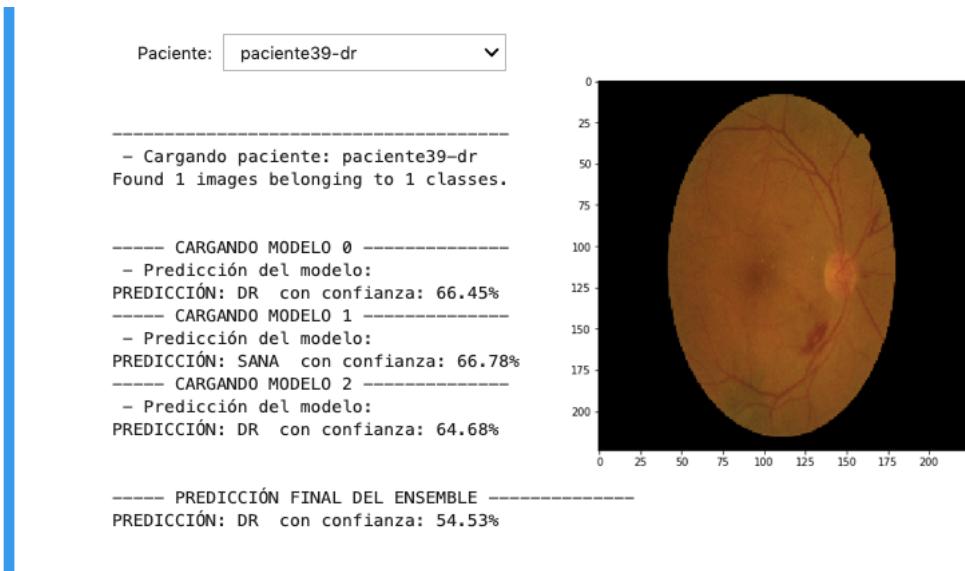


Figura 6.7: Salida del Sistema 3 para una imagen de una retina enferma de RD (omitidos los mapas de activación)

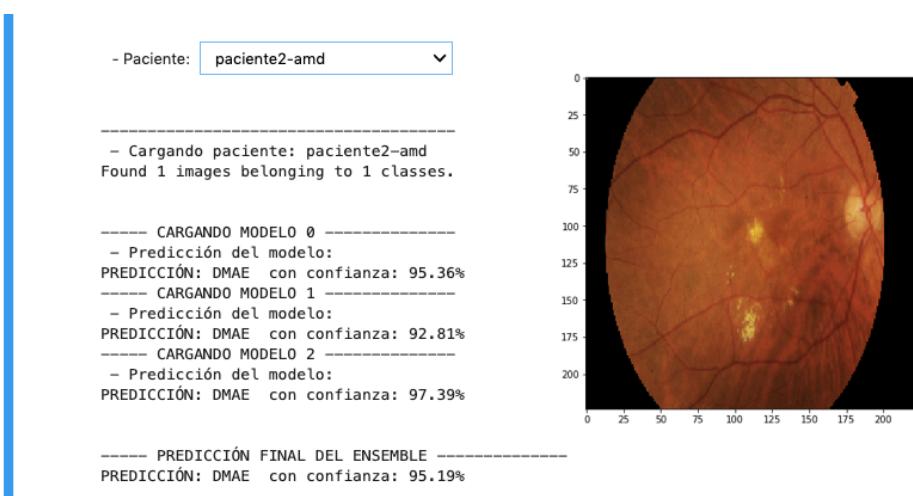


Figura 6.8: Salida del Sistema 3 para una imagen de una retina enferma de DMAE (omitidos los mapas de activación)

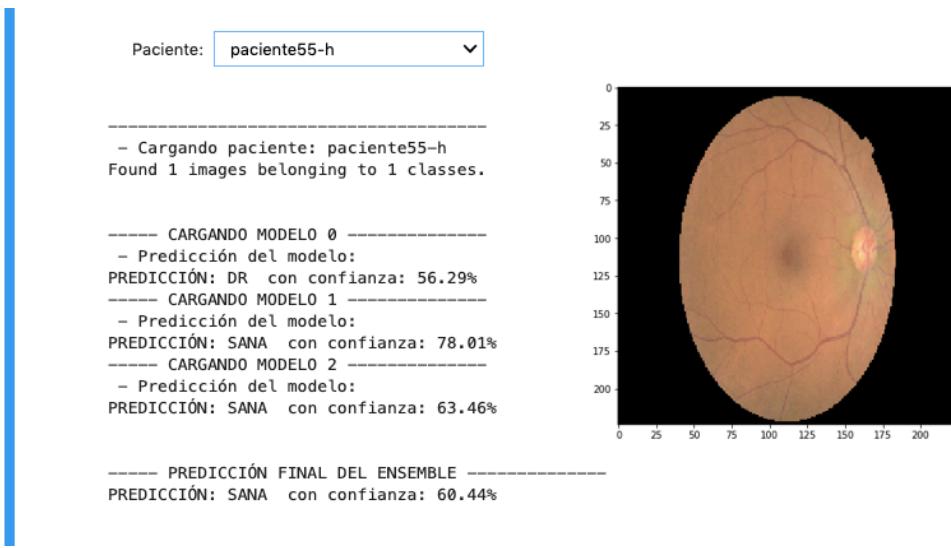


Figura 6.9: Salida del Sistema 3 para una imagen de una retina sana (omitidos los mapas de activación)

6.5. Sistema de Predicción e Interpretación

El Sistema de Predicción e Interpretación permite usar todos los modelos explicados anteriormente simplemente seleccionando una imagen de un paciente.

En las Figuras 6.10 y 6.11 se muestra el Sistema de Predicción devolviendo la predicción realizada por las dos etapas del Sistema 2 para la imagen de un paciente con Degeneración Macular. Previamente, el usuario ha tenido que añadir la imagen de fondo de ojo a una carpeta y haber seleccionado el nombre del fichero en el selector de la parte superior.² En los mapas de atención que devuelve el sistema es posible apreciar cómo los modelos han basado su predicción en la presencia de **drusas**.

De la misma forma, las Figuras 6.12, y 6.13 muestran la respuesta de ese mismo sistema para una retina enferma de RD.

La respuesta del **Ensemble de Clasificadores** para una imagen de una retina enferma de DMAE, como se aprecia en las Figuras 6.14, 6.15 y 6.16,

²Aunque el nombre del fichero contenga la palabra AMD (Age-Related Macular Degeneration, en ningún momento el sistema ha utilizado esa información para realizar su predicción.

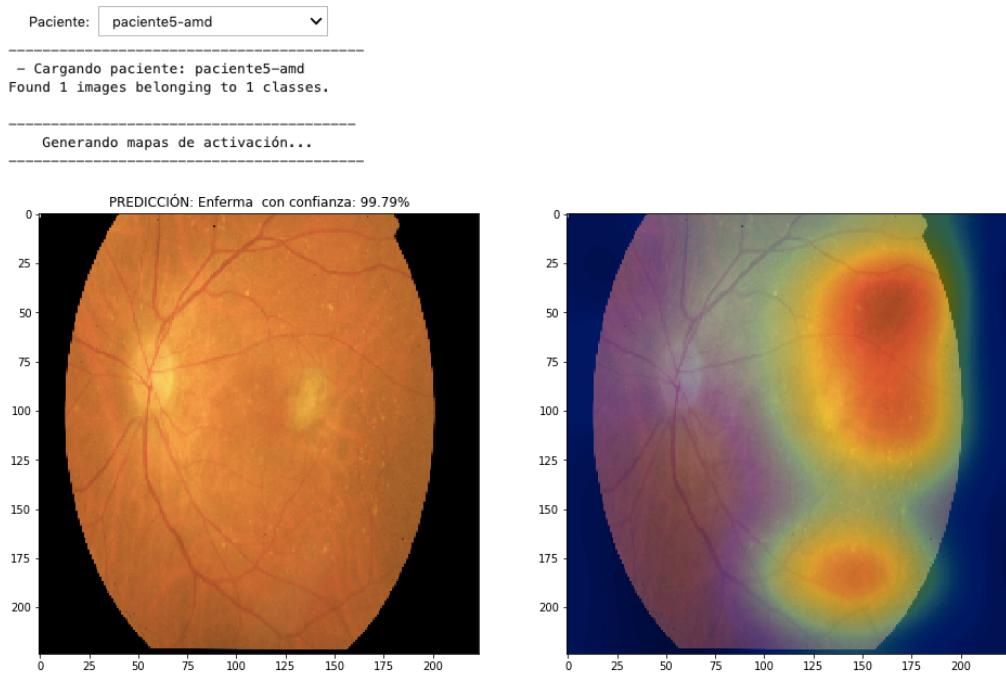


Figura 6.10: Respuesta del Sistema de Predicción a la imagen de una retina con DMAE. Etapa primera del Sistema Multietapa

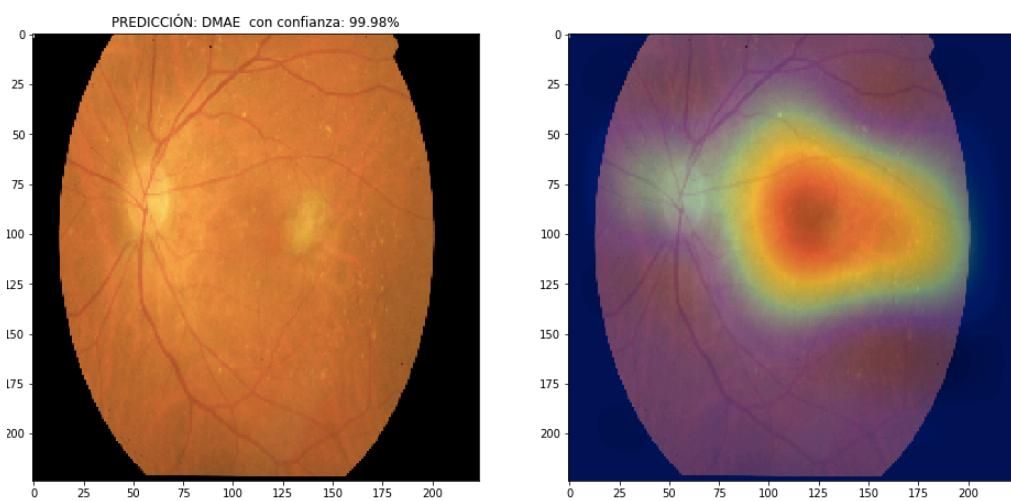


Figura 6.11: Respuesta del Sistema de Predicción a la imagen de una retina con DMAE. Etapa segunda del Sistema Multietapa

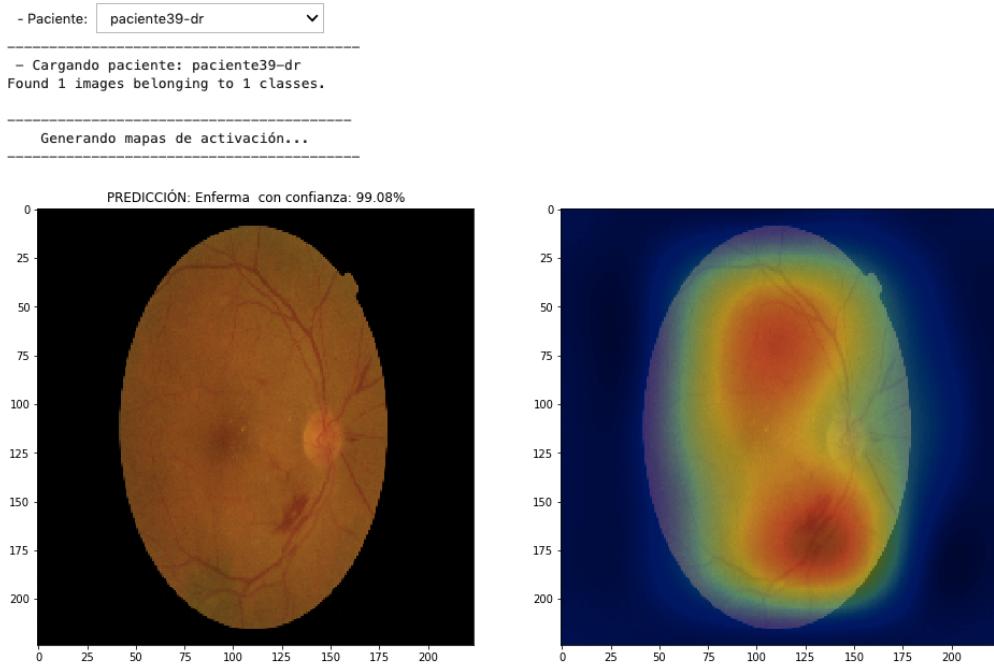


Figura 6.12: Respuesta del Sistema de Predicción a la imagen de una retina con RD. Etapa primera del Sistema Multietapa

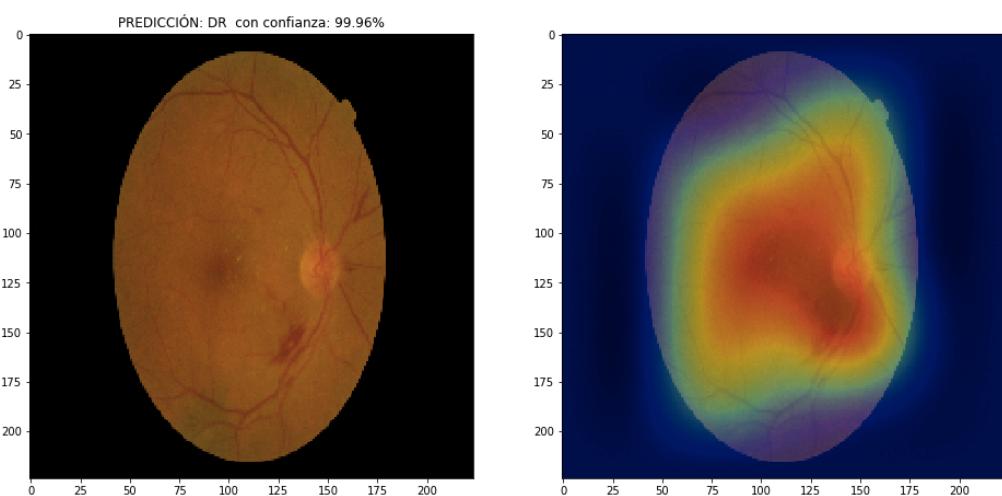


Figura 6.13: Respuesta del Sistema de Predicción a la imagen de una retina con RD. Etapa segunda del Sistema Multietapa

```

- Cargando paciente: paciente7-amd
Found 1 images belonging to 1 classes.
----- CARGANDO MODELO 0 -----
- Predicción del modelo:
PREDICCIÓN: DMAE con confianza: 93.42%
----- CARGANDO MODELO 1 -----
- Predicción del modelo:
PREDICCIÓN: DMAE con confianza: 80.91%
----- CARGANDO MODELO 2 -----
- Predicción del modelo:
PREDICCIÓN: DMAE con confianza: 68.26%

----- PREDICCIÓN FINAL DEL ENSEMBLE -----
PREDICCIÓN: DMAE con confianza: 80.86%

```

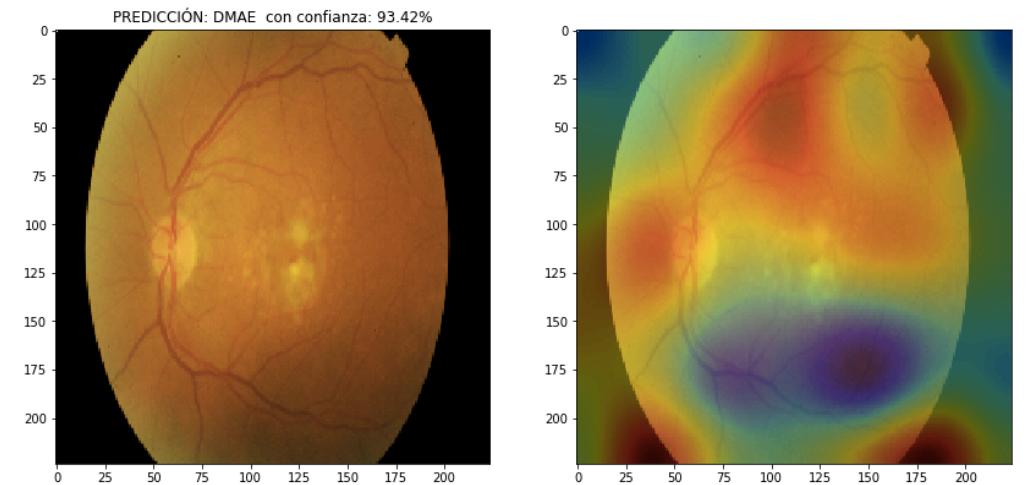


Figura 6.14: Respuesta del Sistema de Predicción a la imagen de una retina con DMAE. Mapa de activación del primer clasificador del Ensemble de Clasificadores

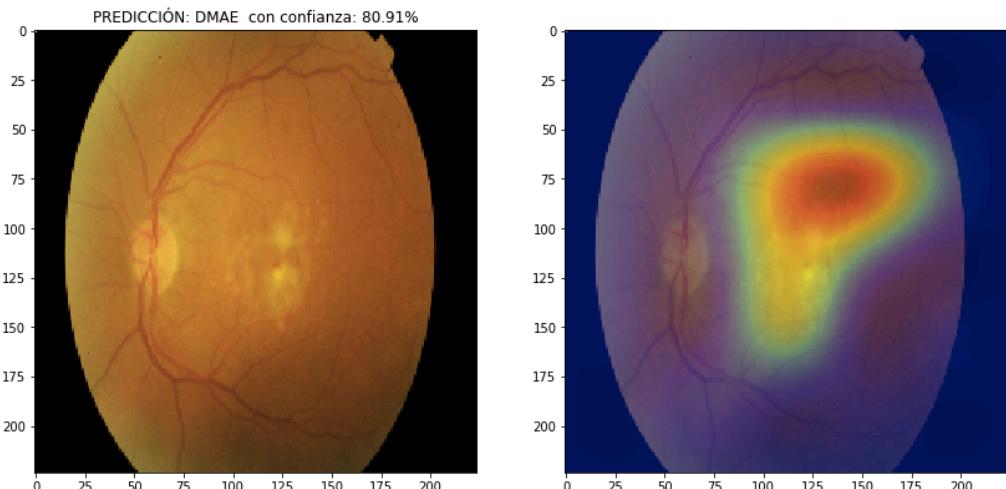


Figura 6.15: Respuesta del Sistema de Predicción a la imagen de una retina con DMAE. Mapa de activación del segundo clasificador del Ensemble de Clasificadores

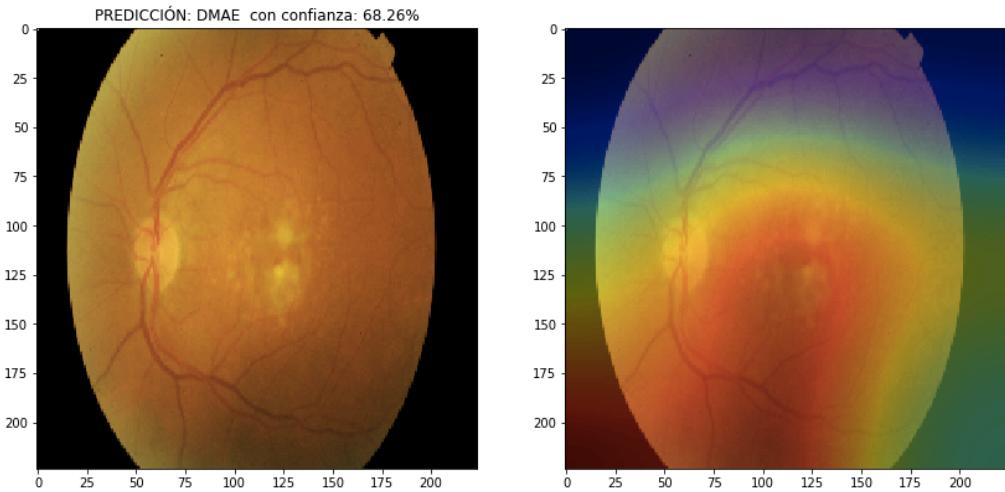


Figura 6.16: Respuesta del Sistema de Predicción a la imagen de una retina con DMAE. Mapa de activación del tercer clasificador del Ensemble de Clasificadores

es muy interesante. Mientras que el primer clasificador basa su clasificación en la posible presencia de neovascularización (Figura 6.14), los otros dos clasificadores (Figuras 6.15 y 6.16) se basan en la presencia de drusas.

Las Figuras 6.17, 6.18 y 6.19 muestran la salida del **Ensemble de Clasificadores** para una imagen de una retina sana. Este caso también es interesante porque ha permitido detectar un **sesgo** en nuestro modelo. La Figura 6.19 muestra cómo este clasificador basa su predicción en la pequeña muesca que existe en la parte superior derecha de la imagen de fondo de ojo. Esto se debe a que, todas las imágenes del dataset de Kaggle presentan esa muesca. Como ese dataset únicamente contiene imágenes de retinas sanas o retinas con RD, nuestro clasificador tiene un sesgo y automáticamente descartará la opción de DMAE cuando vea una imagen de este tipo. Esto puede explicar también el valor tan alto de **accuracy** en la segunda etapa del **Clasificador Multietapa**. Este sesgo deberá ser corregido en posteriores versiones del sistema.

```

- Cargando paciente: paciente69-h
Found 1 images belonging to 1 classes.
----- CARGANDO MODELO 0 -----
- Predicción del modelo:
PREDICCIÓN: SANA con confianza: 57.79%
----- CARGANDO MODELO 1 -----
- Predicción del modelo:
PREDICCIÓN: SANA con confianza: 74.57%
----- CARGANDO MODELO 2 -----
- Predicción del modelo:
PREDICCIÓN: SANA con confianza: 65.97%

----- PREDICCIÓN FINAL DEL ENSEMBLE -----
PREDICCIÓN: SANA con confianza: 66.11%

```

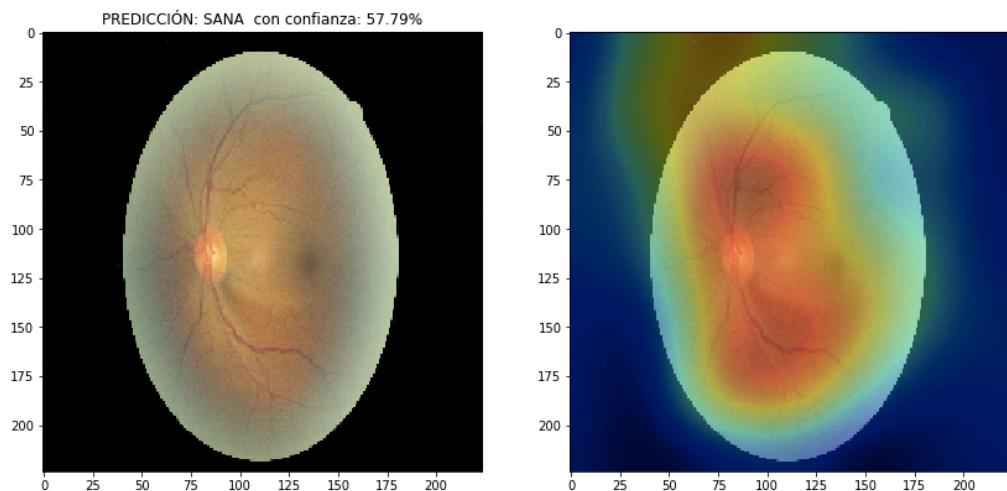


Figura 6.17: Respuesta del Sistema de Predicción a la imagen de una retina sana. Mapa de activación del primer clasificador del Ensemble de Clasificadores

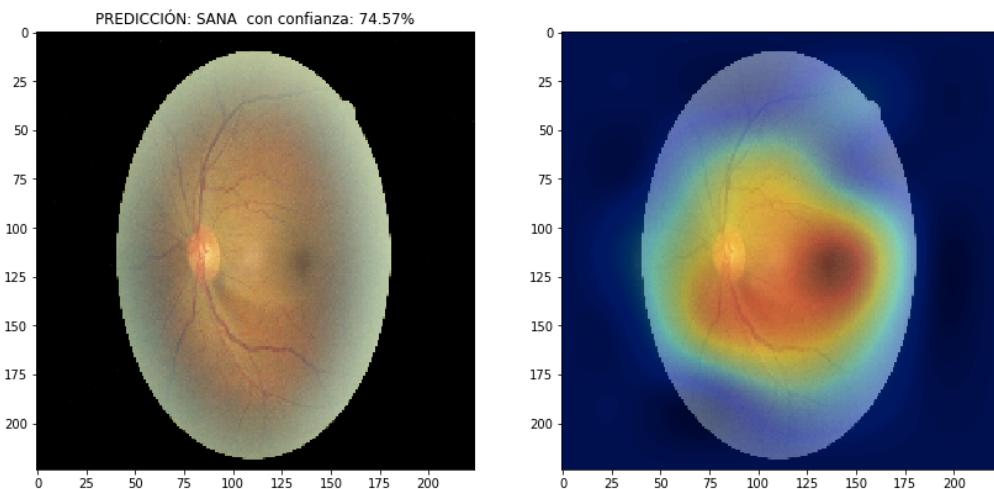


Figura 6.18: Respuesta del Sistema de Predicción a la imagen de una retina sana. Mapa de activación del segundo clasificador del Ensemble de Clasificadores

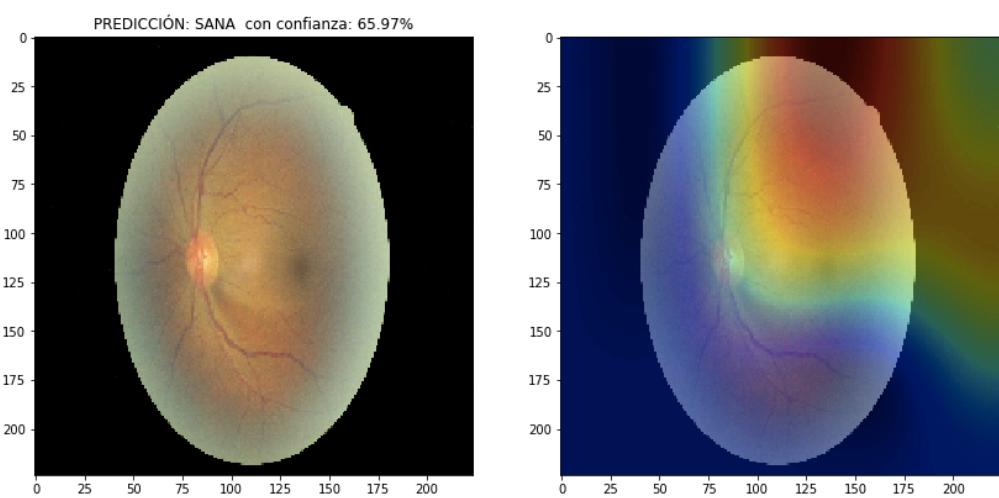


Figura 6.19: Respuesta del Sistema de Predicción a la imagen de una retina sana. Mapa de activación del tercer clasificador del Ensemble de Clasificadores

Capítulo 7

Conclusiones

En plena era de los datos y la automatización, la medicina no puede quedar atrás. Las enfermedades analizadas son solo 2 ejemplos de cómo el Machine Learning puede ayudar a los especialistas a detectar posibles enfermedades en estadios muy tempranos, lo que nos permitirá tratarlas antes de que puedan afectar a la vida diaria del paciente.

Durante este trabajo hemos podido comprobar cómo las dos enfermedades que más casos de ceguera producen en todo el mundo podrían ser detectadas de forma muy temprana, pudiendo así ser tratadas antes de que avancen. El uso de clasificadores de Machine Learning entrenados con datos históricos fiables nos permite crear sistemas robustos que aúnen todo el conocimiento de los mejores expertos y puedan llegar a todos estos sitios donde no es fácil encontrar este tipo de especialistas.

Los resultados obtenidos, incluso estando lejos de los de algunos de los modelos estudiados durante el análisis del estado del arte, son un motivo de optimismo. Como ya se ha explicado a lo largo del trabajo, la mayoría de éstos procedían de datasets con una cantidad muy limitada de imágenes. El sobreajuste, con cantidades tan pequeñas de imágenes es prácticamente inevitable. Los modelos del estado del arte, cuando se utilizaran en *el mundo real* con imágenes procedentes de otras cámaras con distintos tamaños, profundidades de color, artefactos, etc. tendrán serios problemas para generalizar. Sin embargo, nuestro sistema está preparado ante todos esos posibles

cambios gracias a la gran cantidad de datasets utilizados, y al uso del **Data Augmentation**.

Lejos de buscar el *número bonito*, o el *gran titular*, el objetivo de este trabajo, como se puso de manifiesto en el capítulo inicial ha sido siempre crear un sistema verdaderamente **útil** para ser introducido en las clínicas. Para ello, es necesario conseguir un sistema **robusto** e **interpretable**. La robustez nos la proporcionará haber usado más de 39000 imágenes procedentes de 13 conjuntos distintos de datos junto con la combinación de las predicciones de varios modelos con diferentes arquitecturas. La interpretabilidad nos la proporcionará el Sistema de Predicción e Interpretación descrito en apartados anteriores. La información proporcionada por este sistema como las predicciones parciales con su correspondiente confianza de cada clasificador o los mapas de atención ayuda al usuario a entender por qué ha tomado el sistema una decisión concreta y decidir si es fiable la predicción dada. Aunque sea común oír aquello de “*Tortura los datos y te confesarán lo que quieras oír*”, en este caso esa frase no describe la forma de trabajar que ha sido utilizada.

7.1. Trabajo futuro

Una vez creado un sistema inicial verdaderamente útil, robusto y escalable conseguir esos *números bonitos* de los que hablábamos anteriormente, requerirá principalmente de tres elementos: **nuevas imágenes, mayor preprocesamiento, y mayor capacidad de computación para entrenar modelos más complejos**.

El hecho de que el entrenamiento de algunos de los modelos llegara a durar hasta 96 horas, no ha permitido realizar tantas ejecuciones como se hubiera deseado. De haber contado con más tiempo o máquinas más potentes, nuevas arquitecturas como **Xception, DenseNet o MobileNet** hubieran sido evaluadas. Además, soluciones de **Automated Machine Learning** como **AutoKeras¹** hubieran sido de gran utilidad para la obtención de arquitecturas más adecuadas al problema analizado.

¹<https://autokeras.com/>

Aún quedan preguntas en el aire, y no tienen fácil respuesta. Estas preguntas giran alrededor de cómo sería la puesta en producción de este sistema en los servicios de salud de todo el mundo. Este sistema nunca debería ser usado de forma autónoma y, ante la duda, siempre debería prevalecer la opinión del especialista. Sin embargo, esto no quita que su implantación en las consultas como complemento a la opinión del especialista tendría grandes ventajas permitiendo a éste percibir detalles de los que, quizás, en una primera exploración inicial no se había percatado. Sólo después de un tiempo de evaluación de esta forma en consultas, podría empezar a plantearse dotar al sistema de algo más de autonomía, lo que nos permitiría implementarlo en sistemas de salud donde la cantidad de especialistas disponibles es muy limitada.

Durante los próximos años de democratización del Machine Learning, muchos profesionales entenderán que todos estos sistemas no vienen a sustituirlos sino que son una herramienta más de trabajo como lo pueden ser las tan usadas hojas de cálculo. El Machine Learning está muy lejos de sustituir a las personas en ámbitos extremadamente complicados como la medicina. Y hacer una predicción de si esto algún día pasará es poco más que apostar a un número al azar en una ruleta.

Referencias

- Acharya, R. et al., 2008. Application of higher order spectra for the identification of diabetes retinopathy stages. *Journal of Medical Systems*, 32(6), pp.481-488.
- Acharya, U.R. et al., 2017. Automated screening tool for dry and wet age-related macular degeneration (ARMD) using pyramid of histogram of oriented gradients (PHOG) and nonlinear features. *Journal of Computational Science*, 20, pp.41-51.
- Acharya, U.R. et al., 2009. Computer-based detection of diabetes retinopathy stages using digital fundus images. *Proceedings of the institution of mechanical engineers, part H: journal of engineering in medicine*, 223(5), pp.545-553.
- Acharya, U.R. et al., 2012. An integrated index for the identification of diabetic retinopathy stages using texture parameters. *Journal of medical systems*, 36(3), pp.2011-2020.
- Anón, iChallenge AMD Dataset.
- Baldi, P., Sadowski, P. & Whiteson, D., 2014. Searching for exotic particles in high-energy physics with deep learning. *Nature communications*, 5, p.4308.
- Ball, J., Balogh, E. & Miller, B.T., 2015. *Improving diagnosis in health care*, National Academies Press.
- Bjørvig, S., Johansen, M.A. & Fossen, K., 2002. An economic analysis of screening for diabetic retinopathy. *Journal of Telemedicine and Telecare*, 8(1), pp.32-35.
- Burlina, P. et al., 2011. Automatic screening of age-related macular degeneration and retinal abnormalities. En *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, pp. 3962-3966.
- Burlina, P. et al., 2016. Detection of age-related macular degeneration via deep learning. En *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. IEEE, pp. 184-188.
- Cade, W.T., 2008. Diabetes-related microvascular and macrovascular diseases in the physical therapy setting. *Physical therapy*, 88(11), pp.1322-1335.

- Colas, E. et al., 2016. Deep learning approach for diabetic retinopathy screening. *Acta Ophthalmologica*, 94.
- Collobert, R. et al., 2011. Natural language processing (almost) from scratch. *Journal of machine learning research*, 12(Aug), pp.2493-2537.
- Costa, P. & Campilho, A., 2017. Convolutional bag of words for diabetic retinopathy detection from eye fundus images. *IPSJ Transactions on Computer Vision and Applications*, 9(1), p.10.
- Cruz-Roa, A.A. et al., 2013. A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. En *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 403-410.
- Cuadros, J. & Bresnick, G., 2009. EyePACS: an adaptable telemedicine system for diabetic retinopathy screening. *Journal of diabetes science and technology*, 3(3), pp.509-516.
- Currie, J., Lin, W. & Meng, J., 2014. Addressing antibiotic abuse in China: An experimental audit study. *Journal of development economics*, 110, pp.39-51.
- Darwin, C., 2004. *On the origin of species*, 1859, Routledge.
- Decencière, E. et al., 2013. TeleOphta: Machine learning and image processing methods for teleophthalmology. *Irbm*, 34(2), pp.196-203.
- Decencière, E. et al., 2014. Feedback on a publicly distributed image database: the Mesidor database. *Image Analysis & Stereology*, 33(3), pp.231-234.
- De Fauw, J. et al., 2018. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature medicine*, 24(9), p.1342.
- Deng, L., Hinton, G. & Kingsbury, B., 2013. New types of deep neural network learning for speech recognition and related applications: An overview. En *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, pp. 8599-8603.
- Ege, B.M. et al., 2000. Screening for diabetic retinopathy using computer based image analysis and statistical classification. *Computer methods and programs in biomedicine*, 62(3), pp.165-175.
- Englmeier, K. et al., 2004. Early detection of diabetes retinopathy by new algorithms for automatic recognition of vascular changes. *European journal of medical research*, 9(10), pp.473-478.
- Escobar, G.J. et al., 2016. Piloting electronic medical record-based early detection of inpatient deterioration in community hospitals. *Journal of hospital medicine*, 11, pp.S18-S24.

- Farnell, D.J. et al., 2008. Enhancement of blood vessels in digital fundus photographs via the application of multiscale line operators. *Journal of the Franklin institute*, 345(7), pp.748-765.
- Fong, D.S. et al., 2004. Diabetic retinopathy. *Diabetes care*, 27(10), pp.2540-2553.
- Gang, L., Chutatape, O. & Krishnan, S.M., 2002. Detection and measurement of retinal vessels in fundus images using amplitude modified second-order Gaussian filter. *IEEE transactions on Biomedical Engineering*, 49(2), pp.168-172.
- García-Floriano, A. et al., 2017. A machine learning approach to medical image classification: Detecting age-related macular degeneration in fundus images. *Computers & Electrical Engineering*.
- Gargyea, R. & Leng, T., 2017. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*, 124(7), pp.962-969.
- Giancardo, L. et al., 2012. Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. *Medical image analysis*, 16(1), pp.216-226.
- Gondal, W.M. et al., 2017. Weakly-supervised localization of diabetic retinopathy lesions in retinal fundus images. En *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, pp. 2069-2073.
- Goodfellow, I., Bengio, Y. & Courville, A., 2016. *Deep Learning*, MIT Press.
- Grassmann, F. et al., 2018. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. *Ophthalmology*, 125(9), pp.1410-1420.
- Group, A.-R.E.D.S.R. & others, 2001. A randomized, placebo-controlled, clinical trial of high-dose supplementation with vitamins C and E, beta carotene, and zinc for age-related macular degeneration and vision loss: AREDS report no. 8. *Archives of ophthalmology*, 119(10), p.1417.
- Group, E.T.D.R.S.R. & others, 1991. Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified Airlie House classification: ETDRS report number 10. *Ophthalmology*, 98(5), pp.786-806.
- Guariguata, L. et al., 2014. Global estimates of diabetes prevalence for 2013 and projections for 2035. *Diabetes research and clinical practice*, 103(2), pp.137-149.
- Gulshan, V. et al., 2016. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22), pp.2402-2410.
- Hatanaka, Y. et al., 2008. Improvement of automatic hemorrhage detection methods using

brightness correction on fundus images. En *Medical Imaging 2008: Computer-Aided Diagnosis*. International Society for Optics; Photonics, p. 69153E.

Hayashi, J. et al., 2001. A development of computer-aided diagnosis system using fundus images. En *Proceedings Seventh International Conference on Virtual Systems and Multimedia*. IEEE, pp. 429-438.

He, K. et al., 2016. Deep residual learning for image recognition. En *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770-778.

Helmstaedter, M. et al., 2013. Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 500(7461), p.168.

Hijazi, M.H.A., Coenen, F. & Zheng, Y., 2010. Retinal image classification using a histogram based approach. En *The 2010 International Joint Conference on Neural Networks (IJCNN)*. IEEE, pp. 1-7.

Hoover, A., Kouznetsova, V. & Goldbaum, M., 1998. Locating blood vessels in retinal images by piece-wise threshold probing of a matched filter response. En *Proceedings of the AMIA Symposium*. American Medical Informatics Association, p. 931.

Hunter, A. et al., 2000. Quantification of diabetic retinopathy using neural networks and sensitivity analysis. En *Artificial Neural Networks in Medicine and Biology*. Springer, pp. 81-86.

IAPB, 2016. International Agency for the Prevention of Blindness (IAPB). Diabetic Retinopathy.

IDF, 2017. IDF Diabetes Atlas 8th Edition 2017.

Jelinek, H.J. et al., 2006. An automated microaneurysm detector as a tool for identification of diabetic retinopathy in rural optometric practice. *Clinical and Experimental Optometry*, 89(5), pp.299-305.

Jiang, H. et al., 2018. To trust or not to trust a classifier. En *Advances in Neural Information Processing Systems*. pp. 5541-5552.

Kankanhalli, S. et al., 2013. Automated classification of severity of age-related macular degeneration from fundus photographs. *Investigative ophthalmology & visual science*, 54(3), pp.1789-1796.

Katz, N. et al., 1989. Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE Trans. Med. Imaging*, 8(3), pp.263-269.

Kauppi, T. et al., 2006. DIARETDB0: Evaluation database and methodology for diabetic retinopathy algorithms. *Machine Vision and Pattern Recognition Research Group, Lappeenranta University of Technology, Finland*, 73, pp.1-17.

- Klein, R. et al., 1984. The Wisconsin Epidemiologic Study of Diabetic Retinopathy: III. Prevalence and risk of diabetic retinopathy when age at diagnosis is 30 or more years. *Archives of ophthalmology*, 102(4), pp.527-532.
- Krizhevsky, A., Sutskever, I. & Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. En *Advances in neural information processing systems*. pp. 1097-1105.
- LeCun, Y., Bengio, Y. & Hinton, G., 2015. Deep learning. *nature*, 521(7553), p.436.
- LeCun, Y. et al., 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), pp.2278-2324.
- Li, H. & Chutatape, O., 2000. Fundus image features extraction. En *Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Cat. No. 00CH37143)*. IEEE, pp. 3071-3073.
- Li, X. et al., 2017. Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification. En *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE, pp. 1-11.
- Lipton, Z.C., 2016. The mythos of model interpretability. *arXiv preprint arXiv:1606.03490*.
- Lowell, J. et al., 2004. Optic nerve head segmentation. *IEEE Transactions on medical Imaging*, 23(2), pp.256-264.
- Mandel, J.C. et al., 2016. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. *Journal of the American Medical Informatics Association*, 23(5), pp.899-908.
- Maninis, K.-K. et al., 2016. Deep retinal image understanding. En *International conference on medical image computing and computer-assisted intervention*. Springer, pp. 140-148.
- Mansour, R.F., 2018. Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy. *Biomedical engineering letters*, 8(1), pp.41-57.
- Mansour, R.F., 2017. Evolutionary computing enriched computer-aided diagnosis system for diabetic retinopathy: a survey. *IEEE reviews in biomedical engineering*, 10, pp.334-349.
- Mookiah, M.R.K. et al., 2013. Computer-aided diagnosis of diabetic retinopathy: A review. *Computers in biology and medicine*, 43(12), pp.2136-2155.
- Mookiah, M.R.K. et al., 2014. Automated diagnosis of age-related macular degeneration using greyscale features from digital fundus images. *Computers in biology and medicine*, 53, pp.55-64.

- Mookiah, M.R.K. et al., 2014. Decision support system for age-related macular degeneration using discrete wavelet transform. *Medical & biological engineering & computing*, 52(9), pp.781-796.
- Mookiah, M.R.K. et al., 2013. Evolutionary algorithm based classifier parameter tuning for automatic diabetic retinopathy grading: A hybrid feature extraction approach. *Knowledge-based systems*, 39, pp.9-22.
- Niemeijer, M. et al., 2009. Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs. *IEEE transactions on medical imaging*, 29(1), pp.185-195.
- Odstrcilik, J. et al., 2013. Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database. *IET Image Processing*, 7(4), pp.373-383.
- Osareh, A. et al., 2002. Classification and localisation of diabetic-related eye disease. In *European Conference on Computer Vision*. Springer, pp. 502-516.
- Oyster, C.W., 1999. The human eye. *Sunderland, MA: Sinauer*.
- Pan, S.J. & Yang, Q., 2009. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10), pp.1345-1359.
- Pascolini, D. & Mariotti, S.P., 2012. Global estimates of visual impairment: 2010. *British Journal of Ophthalmology*, 96(5), pp.614-618.
- Pead, E. et al., 2019. Automated detection of age-related macular degeneration in color fundus photography: a systematic review. *survey of ophthalmology*, 64(4), pp.498-511.
- Phan, T.V. et al., 2016. Automatic screening and grading of age-related macular degeneration from texture analysis of fundus images. *Journal of ophthalmology*, 2016.
- Pratt, H. et al., 2016. Convolutional neural networks for diabetic retinopathy. *Procedia Computer Science*, 90, pp.200-205.
- Quellec, G. et al., 2017. Deep image mining for diabetic retinopathy screening. *Medical image analysis*, 39, pp.178-193.
- Quellec, G. et al., 2008. Optimal wavelet transform for the detection of microaneurysms in retina photographs. *IEEE transactions on medical imaging*, 27(9), pp.1230-1241.
- Rajkomar, A., Dean, J. & Kohane, I., 2019. Machine learning in medicine. *New England Journal of Medicine*, 380(14), pp.1347-1358.
- Rajkomar, A. et al., 2018. Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine*, 1(1), p.18.

- Reza, A.W. & Eswaran, C., 2011. A decision support system for automatic screening of non-proliferative diabetic retinopathy. *Journal of medical systems*, 35(1), pp.17-24.
- Ruamviboonsuk, P. et al., 2005. Screening for diabetic retinopathy in rural area using single-field, digital fundus images. *J Med Assoc Thai*, 88(2), pp.176-180.
- Schwartz, W.B., 1970. Medicine and the computer: the promise and problems of change. In *Use and Impact of Computers in Clinical Medicine*. Springer, pp. 321-335.
- Schwartz, W.B., Patil, R.S. & Szolovits, P., 1986. Artificial Intelligence in Medicine Where Do We Stand. *Jurimetrics J.*, 27, p.362.
- Sellahewa, L. et al., 2014. Grader agreement, and sensitivity and specificity of digital photography in a community optometry-based diabetic eye screening program. *Clinical Ophthalmology (Auckland, NZ)*, 8, p.1345.
- Selvaraju, R.R. et al., 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision*. pp. 618-626.
- Simonyan, K. & Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Sinthanayothin, C. et al., 2002. Automated detection of diabetic retinopathy on digital fundus images. *Diabetic medicine*, 19(2), pp.105-112.
- Sinthanayothin, C. et al., 2003. Automated screening system for diabetic retinopathy. In *3rd International Symposium on Image and Signal Processing and Analysis, 2003. ISPA 2003. Proceedings of the*. IEEE, pp. 915-920.
- Slack, W.V. et al., 1966. A computer-based medical-history system. *New England Journal of Medicine*, 274(4), pp.194-198.
- Szegedy, C. et al., 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 1-9.
- Szegedy, C. et al., 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 2818-2826.
- Tan, J.H. et al., 2018. Age-related macular degeneration detection using deep convolutional neural network. *Future Generation Computer Systems*, 87, pp.127-135.
- Tolias, Y.A. & Panas, S.M., 1998. A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustering. *IEEE Transactions on Medical Imaging*, 17(2), pp.263-273.
- Vlachos, M. & Dermatas, E., 2010. Multi-scale retinal vessel segmentation using line

- tracking. *Computerized Medical Imaging and Graphics*, 34(3), pp.213-227.
- Walter, T. et al., 2007. Automatic detection of microaneurysms in color fundus images. *Medical image analysis*, 11(6), pp.555-566.
- Wang, H. et al., 2000. An effective approach to detect lesions in color retinal images. En *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*. IEEE, pp. 181-186.
- WHO & others, 2013. Universal eye health: a global action plan 2014-2019.
- Williams, R. et al., 2004. Epidemiology of diabetic retinopathy and macular oedema: a systematic review. *Eye*, 18(10), p.963.
- Wong, W.L. et al., 2014. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *The Lancet Global Health*, 2(2), pp.e106-e116.
- Zhang, P. et al., 2009. Economic impact of diabetes. *Diabetes Atlas, IDF*, 4.
- Zhang, Q.-s. & Zhu, S.-C., 2018. Visual interpretability for deep learning: a survey. *Frontiers of Information Technology & Electronic Engineering*, 19(1), pp.27-39.
- Zhang, X. et al., 2017. Direct medical cost associated with diabetic retinopathy severity in type 2 diabetes in Singapore. *PloS one*, 12(7), p.e0180949.
- Zheng, Y., He, M. & Congdon, N., 2012. The worldwide epidemic of diabetic retinopathy. *Indian journal of ophthalmology*, 60(5), p.428.
- Zheng, Y., Hijazi, M.H.A. & Coenen, F., 2012. Automated «disease/no disease» grading of age-related macular degeneration by an image mining approach. *Investigative ophthalmology & visual science*, 53(13), pp.8310-8318.
- Zheng, Y., Hijazi, M.H.A. & Coenen, F., 2011. Automated Grading of Age-Related Macular Degeneration by an Image Mining Approach. *Investigative Ophthalmology & Visual Science*, 52(14), pp.6568-6568.
- Zhu, J., Zhang, E. & Del Rio-Tsonis, K., 2001. Eye Anatomy. e LS.