

# 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

Trabajo Final de Máster  
Máster Universitario en Ingeniería de Telecomunicaciones

Tutorizado por:  
Naranjo Orredo, Valery  
Colomer Granero, Adrián



Escuela Técnica Superior de Ingenieros de Telecomunicación  
Universidad Politécnica de Valencia  
Valencia, España

Septiembre 2019

*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

# **Lista de figuras**



# **Lista de tablas**



# Abreviaciones

<b>RD</b>	Retinopatía Diabética
<b>DMAE</b>	Degeneración Macular Asociada a la Edad
<b>NPDR</b>	Non-Proliferative Diabetic Retinopathy
<b>PDR</b>	Proliferative Diabetic Retinopathy
<b>ARIA</b>	Automated Retinal Image Analyzer
<b>IA</b>	Inteligencia Artificial
<b>CNN</b>	Convolutional Neural Network
<b>SVM</b>	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 informe<sup>1</sup> 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.<sup>2</sup> 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.

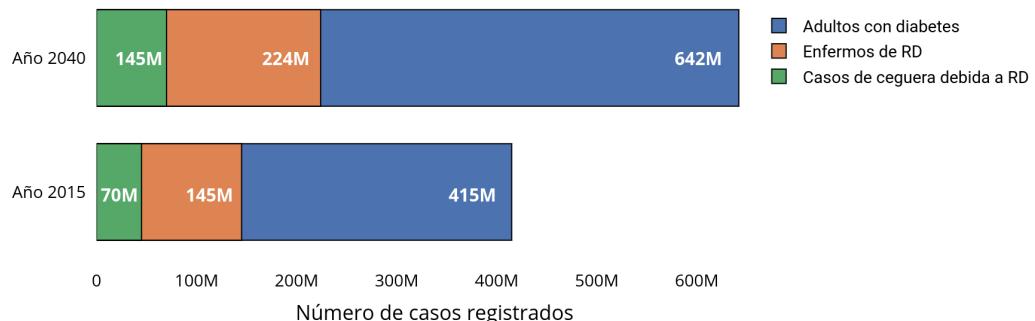


Figura 1.1: Prevalecencia y previsión de crecimiento de la diabetes y la RD a nivel mundial. Gráfico de elaboración propia

La diabetes supone, aproximadamente, el 11.6% del presupuesto total de salud

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

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

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, 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 desconoce 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



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

tipo de patologías.<sup>3</sup>

---

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

## 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**.

- 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,

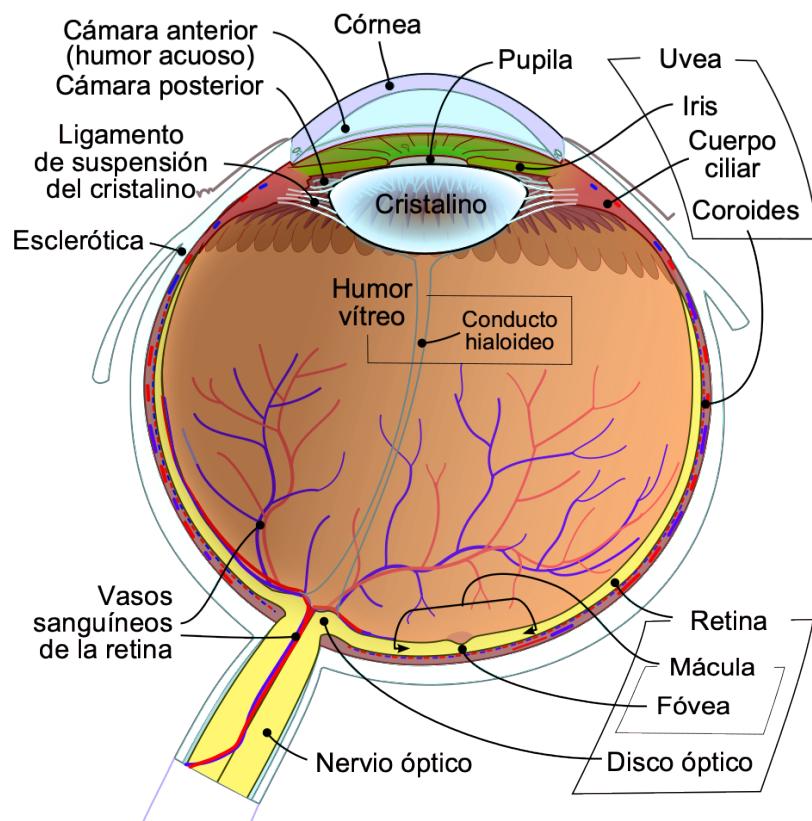


Figura 2.1: Estructura del ojo humano. Fuente: Wikipedia

para poder permitir el paso de la mayor cantidad posible de luz.

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

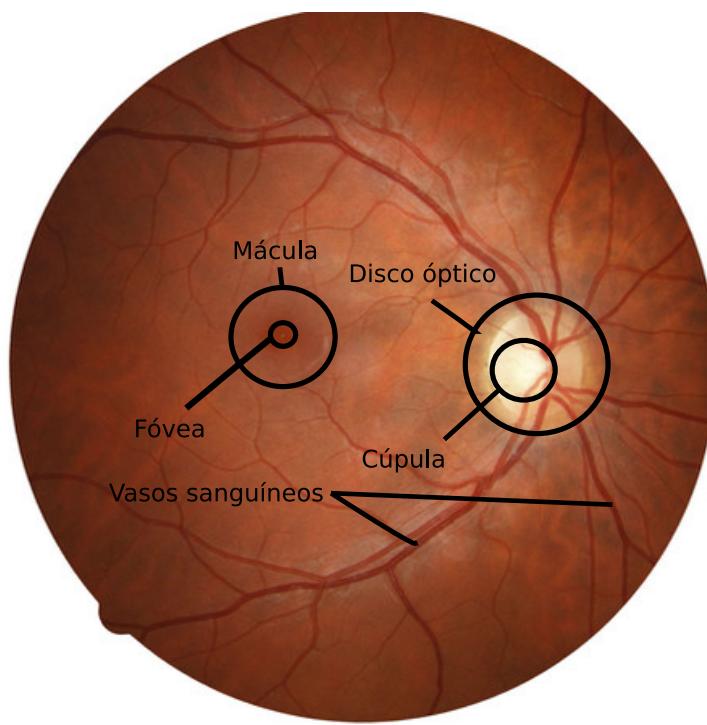


Figura 2.2: Elementos de la retina. Fuente: Kaggle (anotaciones de elaboración propia)

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.

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**

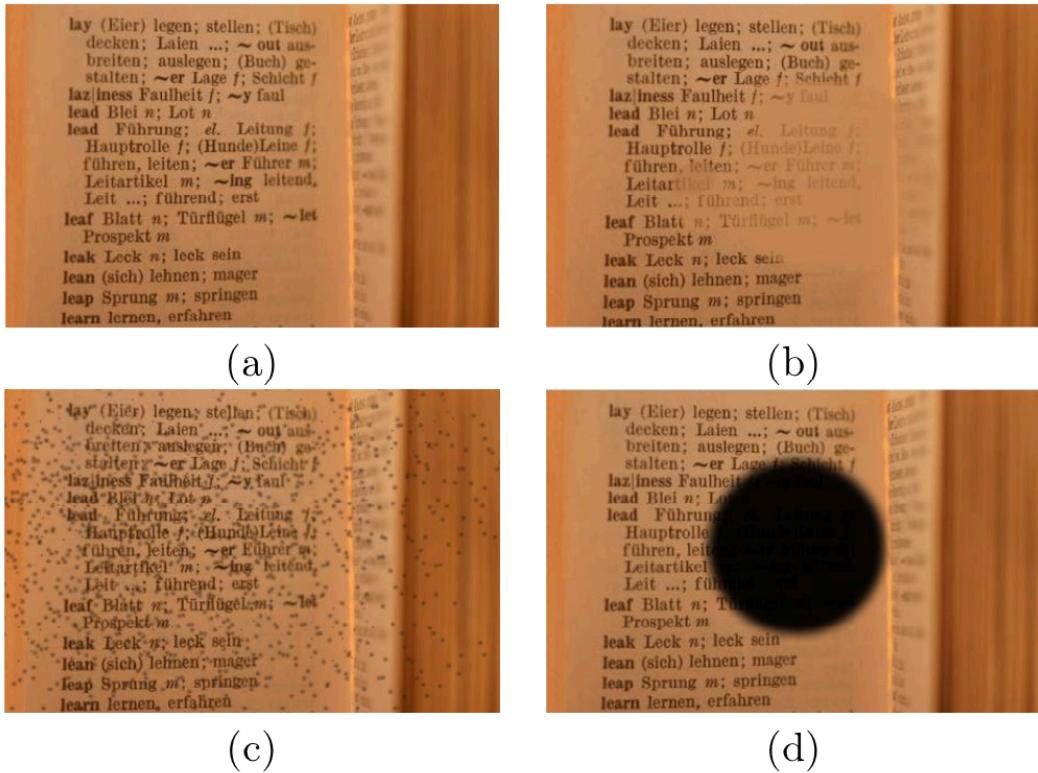


Figura 2.3: Efectos en la visión de las enfermedades analizadas en este trabajo: (a) Visión normal, (b) con Retinopatía Diabética no proliferativa, (c) con Retinopatía Diabética proliferativa y (d) con degeneración macular asociada a la edad. Fuente: American Academy of Ophthalmology ([www.aao.org](http://www.aao.org))

**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.

### 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 con láser**. Este tratamiento se puede realizar en una o varias sesiones, tras haber comprobado, mediante una angiografía fluoresceína 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 má-

genes 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.

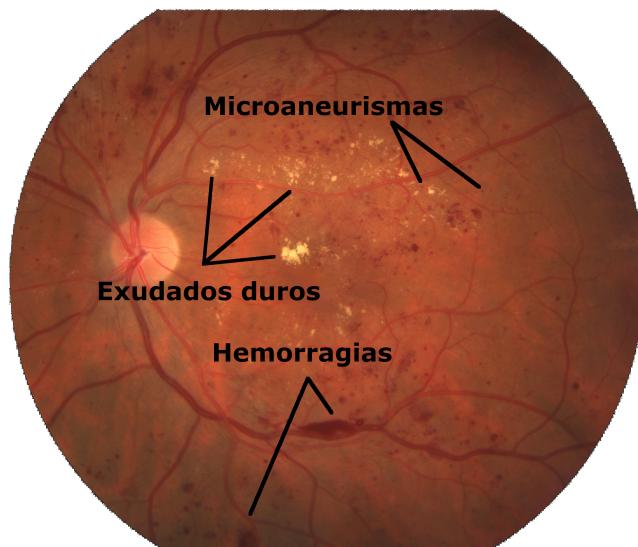


Figura 2.4: Lesiones típicas de la Retinopatía Diabética. Elaboración proia

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 la Retinopatía Diabética en función de las lesiones presentes en el paciente.<sup>1</sup> 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).

---

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

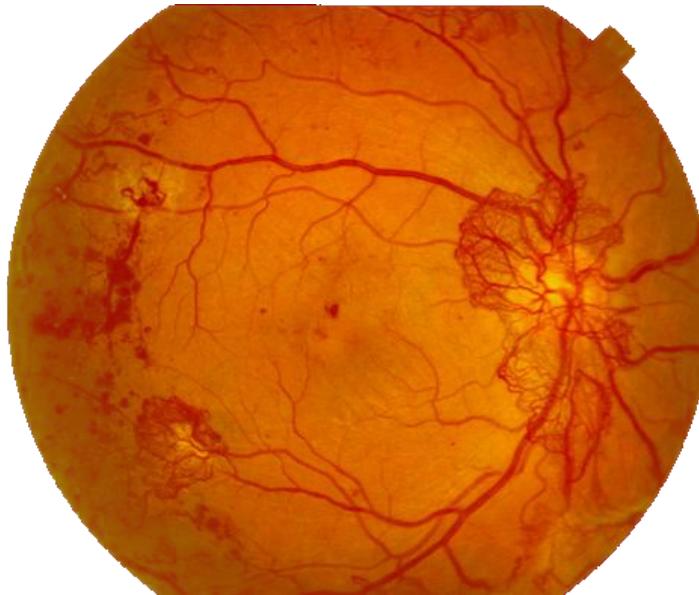


Figura 2.5: Ejemplo de retina en la que se ha producido neovascularización

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

Nivel de gravedad	Observaciones
<b>Grado 0:</b> No RD	Sin ninguna anomalía
<b>Grado 1:</b> NPDR Ligera	Presencia de algunos microaneurismas
<b>Grado 2:</b> NPDR Moderada	Presencia de más microaneurismas pero menos que en el grado 3
<b>Grado 3:</b> NPDR Severa	Alguno de los siguientes: <ul style="list-style-type: none"><li>- Más de 20 hemorragias intrarretinales</li><li>- Dilataciones venosas</li><li>- Anomalías microvasculares en la retina</li><li>- Ningún signo de PDR</li></ul>
<b>Grado 4:</b> PDR	Alguno de los siguientes: <ul style="list-style-type: none"><li>- Neovascularización</li><li>- Hemorragia vítrea</li></ul>

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

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:



Figura 2.6: Retina con drusas a causa de DMAE

- **Estadio inicial:** En este estadio existe un reducido número de drusas redondas y de pequeño tamaño (menos de  $125 \mu\text{m}$ ). 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  $125 \mu\text{m}$ . 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.

### 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 (Ruamviboonsuk 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 notablemen-

te. 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.<sup>1</sup>

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:

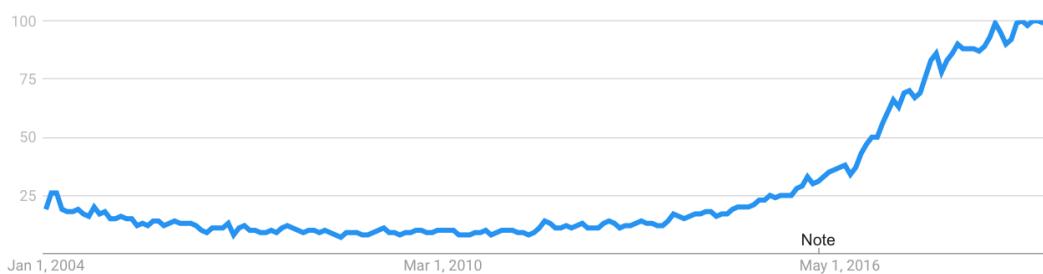


Figura 3.1: Interés, a lo largo del tiempo y en todo el mundo, del término Machine Learning en el buscador Google. Datos de Enero de 2014 a Julio de 2019. Un valor de 100 indica la máxima popularidad del término. Los valores 50 y 0 indican, respectivamente, que un término es la mitad de popular en relación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

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

- **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 en 2018 en todo el mundo, se generaron entre 2016 y 2017.<sup>2</sup>

### 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.

---

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

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 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.

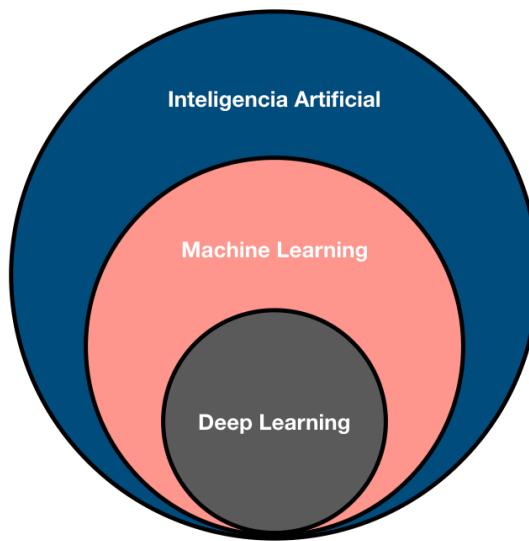


Figura 3.2: El Machine Learning es un campo perteneciente a la Inteligencia Artificial. El Deep Learning, a su vez, es un campo dentro del Machine Learning. Elaboración propia

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 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.

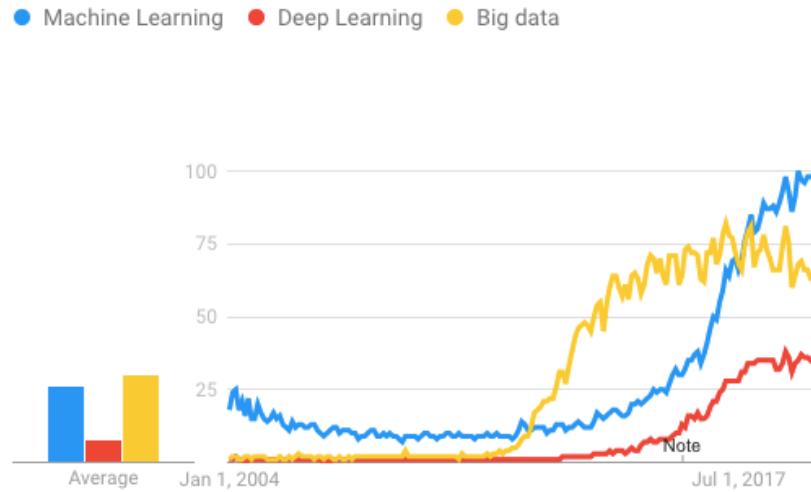


Figura 3.3: Interés, a lo largo del tiempo y en todo el mundo, de los términos Machine Learning (en azul), Deep Learning (en rojo) y Big Data (en amarillo) en el buscador Google. Datos de Enero de 2014 a Julio de 2019. Un valor de 100 indica la máxima popularidad del término. Los valores 50 y 0 indican, 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 backpropagation

Una red neuronal consiste en un conjunto de nodos, conocidos como **neuronas**, conectados entre sí para transmitirse 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

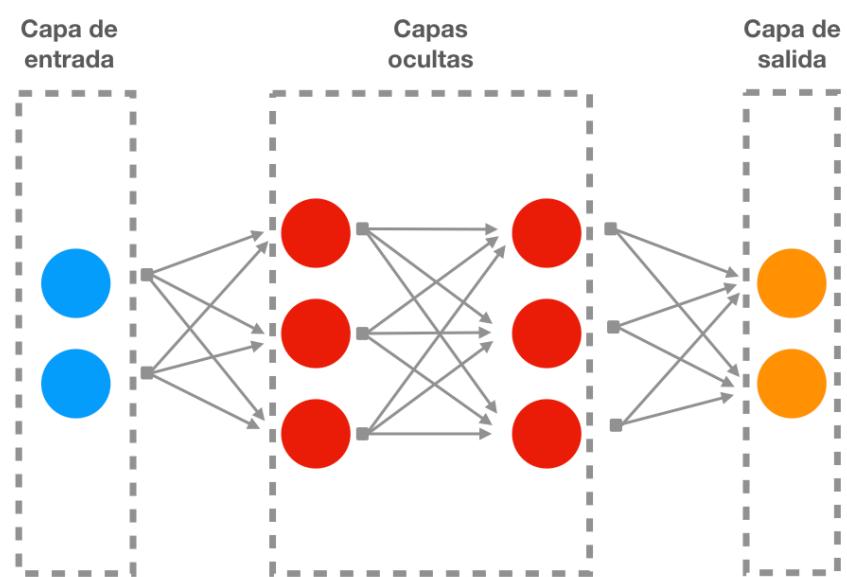


Figura 3.4: Representación de una red neuronal con dos capas ocultas. Cada uno de los círculos representa una neurona. Elaboración propia

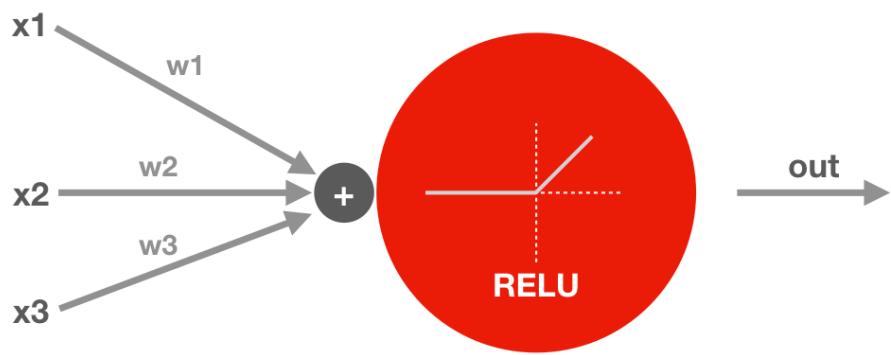


Figura 3.5: Representación de una sola neurona con 3 entradas. Cada una de esas entradas tiene asociado un peso. La neurona utiliza la función de activación ReLU. Elaboración propia

pueden ser positivos o negativos.<sup>3</sup> 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, los resultados de todos estos productos se suman (ecuación 3.2) y se pasa el valor obtenido a lo que se conoce como **función de activación** (ecuación 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)<sup>4</sup> 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.

---

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

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

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

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 en el sentido opuesto al vector de gradiente (ecuación 3.6), buscando así **minimizar el error**. El valor  $\alpha$  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)**<sup>5</sup> 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.

---

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

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.

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 convolución<sup>6</sup> 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.

---

<sup>6</sup>Aunque 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)

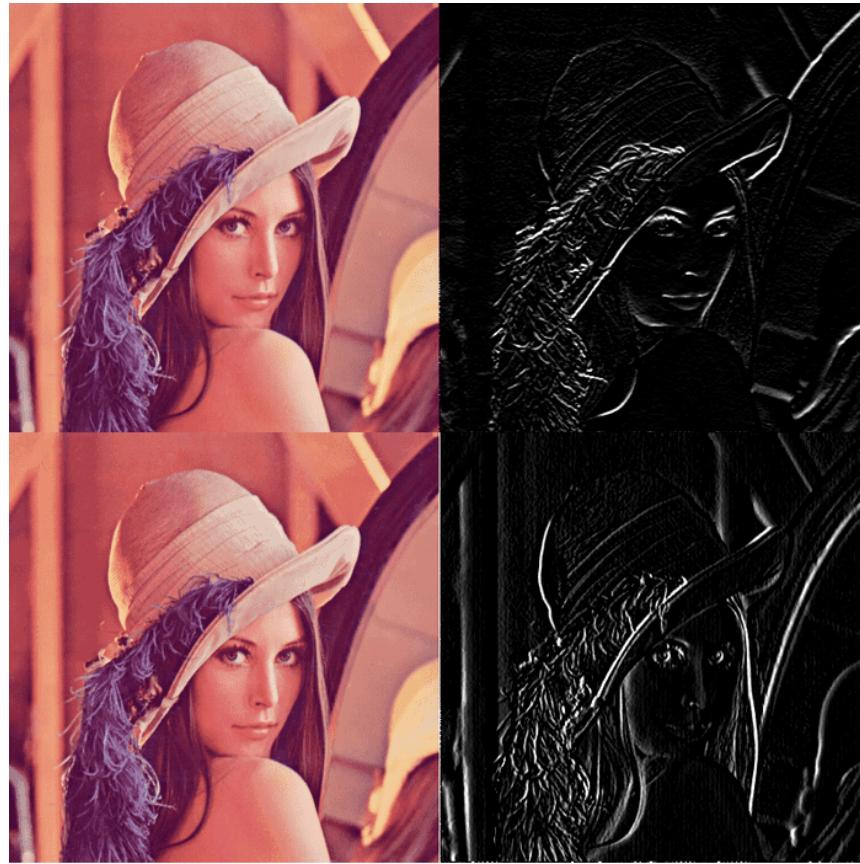


Figura 3.6: Resultado de la convolución de una imagen con un filtro Sobel de 3x3 horizontal (arriba) y otro vertical (abajo) Fuente: <https://victorzhou.com/blog/intro-to-cnns-part-1/>

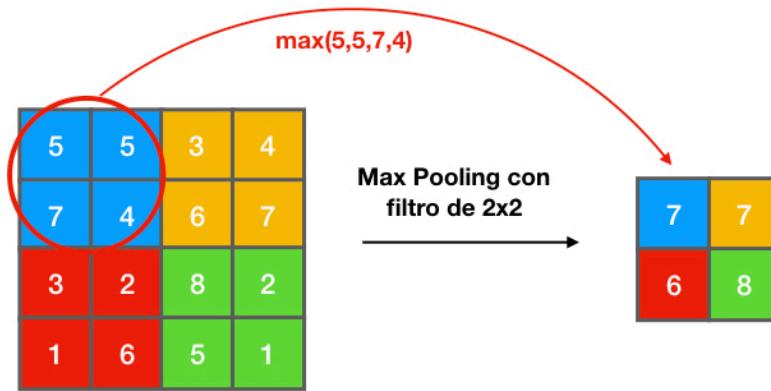


Figura 3.7: Representación del proceso de Max Pooling con un filtro de 2x2 sobre una imagen de 4x4. Elaboración propia

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.

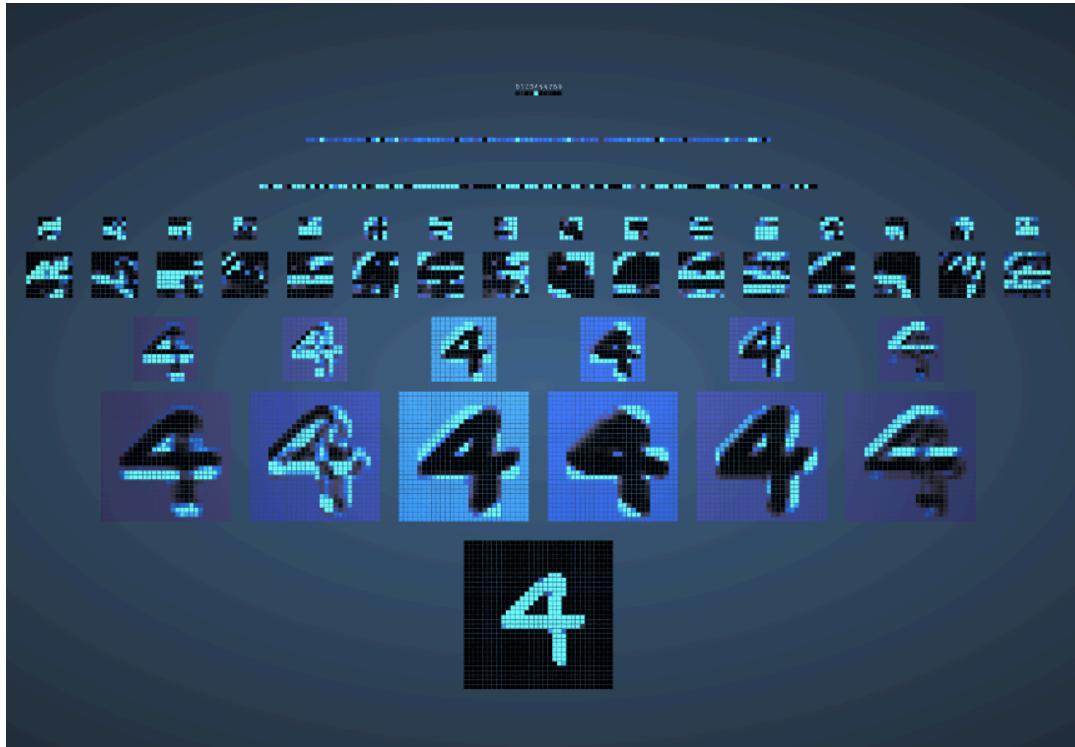


Figura 3.8: Representación de los mapas de activación de una red convolucional con 2 capas convolucionales y 3 fully-connected. Cada capa convolucional va seguida de una de max pooling.  
Fuente: <http://scs.ryerson.ca/~aharley/vis/conv/flat.html>

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 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.<sup>7</sup>

#### 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:

- 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.

---

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

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:<sup>8</sup>

- **El nuevo dataset es pequeño pero similar al original:** Al tratarse

---

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

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.

### 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).

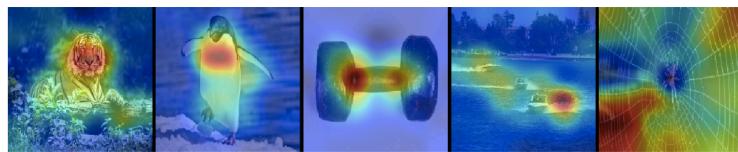


Figura 3.9: Mapas de atención generados por Grad-Cam para distintas clases de Imagenet. Fuente: <https://github.com/raghakot/keras-vis>

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.

### 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.<sup>9</sup>

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 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

---

<sup>9</sup>Aunque no hay que olvidar las limitaciones derivadas de la privacidad y la protección de datos

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.

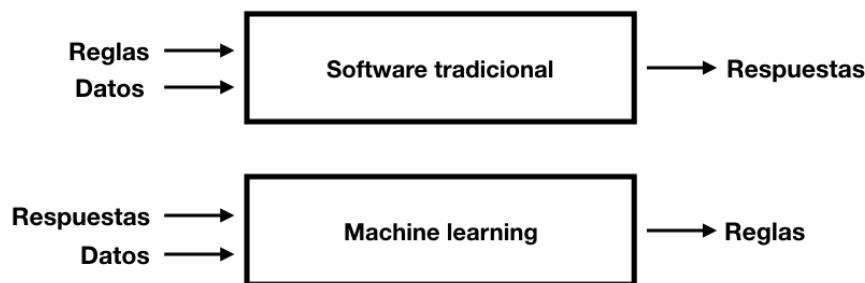


Figura 3.10: Diferencias entre Software tradicional y Machine learning. Elaboración propia.

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 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 posi-

bles 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**.<sup>10</sup> 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).

---

<sup>10</sup><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.

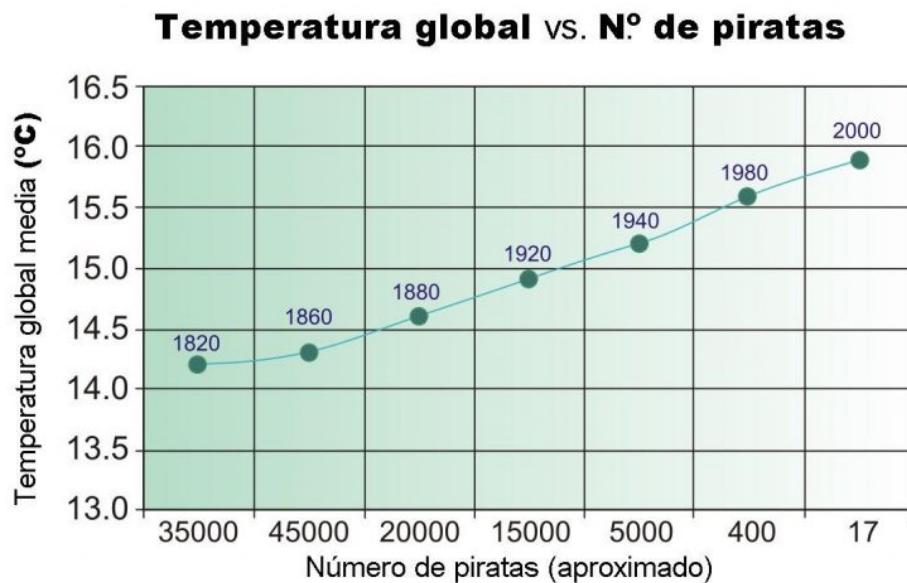


Figura 3.11: Correlación entre el aumento de la temperatura media global y el descenso del número de piratas. Fuente: <https://www.jotdown.es/2016/06/correlacion-no-implica-causalidad/>

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.

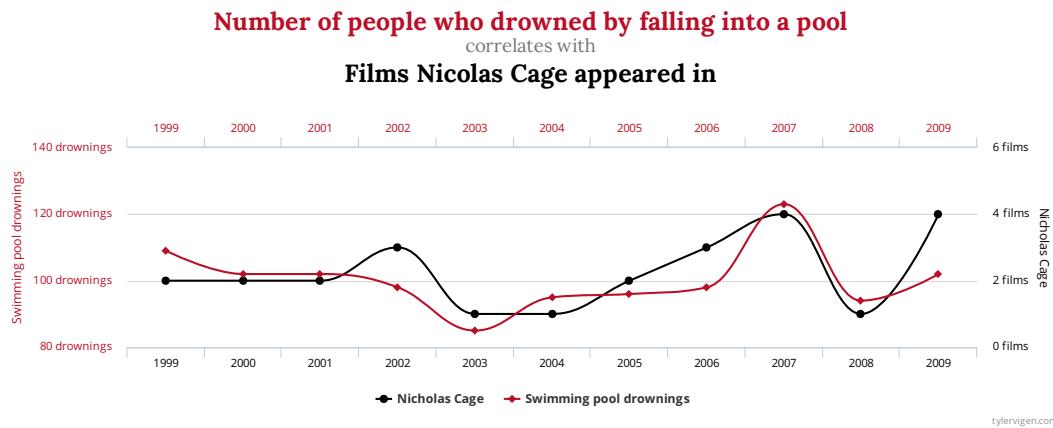


Figura 3.12: Correlación entre el número de ahogados en piscinas de Estados Unidos y el número de apariciones en películas de Nicholas Cage. Fuente: <http://www.tylervigen.com/spurious-correlations>

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 Superviviente<sup>11</sup> y la Paradoja de Simpson<sup>12</sup>).

<sup>11</sup>[https://es.wikipedia.org/wiki/Sesgo\\_del\\_superviviente](https://es.wikipedia.org/wiki/Sesgo_del_superviviente)

<sup>12</sup>[https://es.wikipedia.org/wiki/Paradoja\\_de\\_Simpson](https://es.wikipedia.org/wiki/Paradoja_de_Simpson)

## 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 ??). 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