Universidad de Buenos Aires

Facultad de Ciencias Exactas y Naturales

Maestría en Explotación de Datos y Descubrimiento de Conocimiento

Detección de cataratas y glaucoma en imágenes de fondo de ojo

Hernán Estrin

Julio 2021

Índice de contenidos

1. Introducción

1.1. Enfermedades del ojo humano y su detección

Entre las diversas enfermedades que afectan el ojo humano las cataratas y el glaucoma se ubican como las primeras 2 causas a nivel mundial de ceguera.

Las cataratas se desarrollan como consecuencia de la paulatina opacidad de la transparencia del lente natural del ojo humano, llamado cristalino. Este se encuentra ubicado detrás del iris (la parte coloreada de la pupila) y es el encargado de refractar la luz que ingresa al globo ocular, a través de la cual las imágenes se forman nítidamente en la retina, la membrana ubicada en el fondo del ojo encargada de transformar estas impresiones luminosas en impulsos nerviosos.

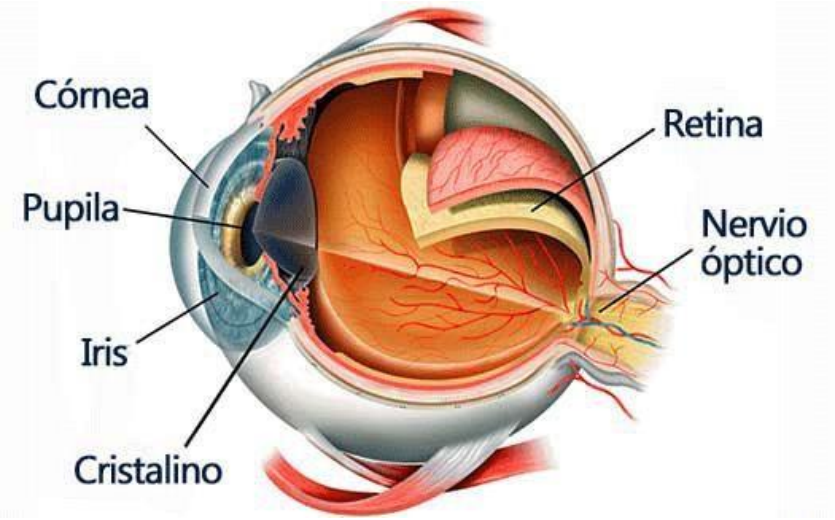


Figura 1. Estructura del ojo humano con sus partes más relevantes detalladas.

Varios factores concomitantes, principalmente la edad, pero también la herencia, la medicación, la diabetes y el medio ambiente hacen que las proteínas que forman el cristalino se comiencen a aglutinar, haciendo que se vuelva más grueso, menos flexible y con ello menos transparente. Esto da lugar al desarrollo de sus síntomas característicos, como la visión borrosa, la pérdida de percepción de los colores y la necesidad de una mayor iluminación para ver con claridad. De acuerdo con las estimaciones de la OMS, alrededor de 100 millones de personas en el mundo padecen cataratas y es la causa más frecuente de ceguera prevenible, aumentando su prevalencia considerablemente después de los 60 años. Si bien los tratamientos actuales, particularmente la cirugía, resultan eficaces en su tratamiento, año tras año millones de personas, en su mayoría de países subdesarrollados, quedan ciegas a causa de esta enfermedad.

Por su parte, el glaucoma es un grupo de afecciones oculares que dañan lenta pero sostenidamente el nervio óptico. Formado por más de un millón de microscópicas fibras nerviosas, este nervio es el responsable de transmitir los impulsos eléctricos generados por la retina hacia el cerebro, donde se completa el proceso de formación de la imagen.

Producido por la excesiva presión dentro del globo ocular, que se genera a su vez como consecuencia de la acumulación de fluido en la parte delantera del ojo, se trata de una enfermedad que en su versión más común es asintomática hasta fases avanzadas, afectando en forma irreversible primero a la visión periférica, ganándose así el título de "ceguera silenciosa". Debido a esta condición cobra particular relevancia su detección temprana mediante un examen oftalmológico completo, sobre todo en individuos con mayor propensión a desarrollarlo, como aquellos mayores de 60 años, con antecedentes familiares, diabetes y/o presión arterial alta. En la actualidad alrededor de 80 millones de personas la padecen a nivel mundial, siendo la medicación y cirugía laser eficientes para detener su avance, pero no para recuperar la visión perdida.

Para la detección de estas y otras patologías graves, el examen de fondo de ojo, técnicamente conocida como oftalmoscopia, es muy utilizada. Se trata de una técnica indolora que permite tomar una imagen unidimensional de las estructuras oculares, a través de la pupila, y tal y como se hallan dentro del ojo. En el caso de las cataratas, la retina se suele observar borrosa precisamente porque la luz no es refractada correctamente; en el caso del glaucoma, este método permite detectar cambios físicos característicos en la cabeza del nervio óptico, llevándose a cabo pruebas complementarias *a posteriori* para confirmar el diagnóstico.

1.2. Redes Neuronales Convolucionales

Dentro del universo del Aprendizaje Profundo*,* las redes neuronales convolucionales, conocida por sus siglas en inglés *CNN,* son un tipo de redes neuronales que se especializa en trabajar con datos dispuestos en forma de grilla, es decir datos en los cuales la relación espacial entre las distintas características o *features* de los datos con que se cuenta son relevantes para resolver el problema deseado.

Las imágenes, dispuestas como píxeles ordenados en uno o más canales, son el ejemplo más representativo de este arreglo de datos, por lo que el uso de *CNNs* para resolver tareas de aprendizaje automático está ampliamente difundido y representa, a la fecha, el estado del arte. Entre las principales problemáticas para las cuales se han utilizado con este tipo de datos se cuentan, entre otros, la detección de objetos, y la segmentación y clasificación de imágenes.

Inspiradas en el estudio de la corteza visual del cerebro humano, su estructura básica consiste en varias capas de filtros convolucionales que se aplican sucesivamente sobre las imágenes de entrada, cuya tarea consiste en crear representaciones de las mismas que resulten útiles para resolver el problema propuesto. Equiparables a los patrones que los humanos utilizan para

reconocer algo en una imagen, estos filtros son matrices numéricas que van “recorriendo” la imagen que reciben como entrada, resaltando en su salida distintas características de las mismas, con la particularidad que los pesos o parámetros que las componen son aprendidos por la red de modo de obtener el mejor resultado posible, valiéndose del conocido mecanismo de backpropagation[i].

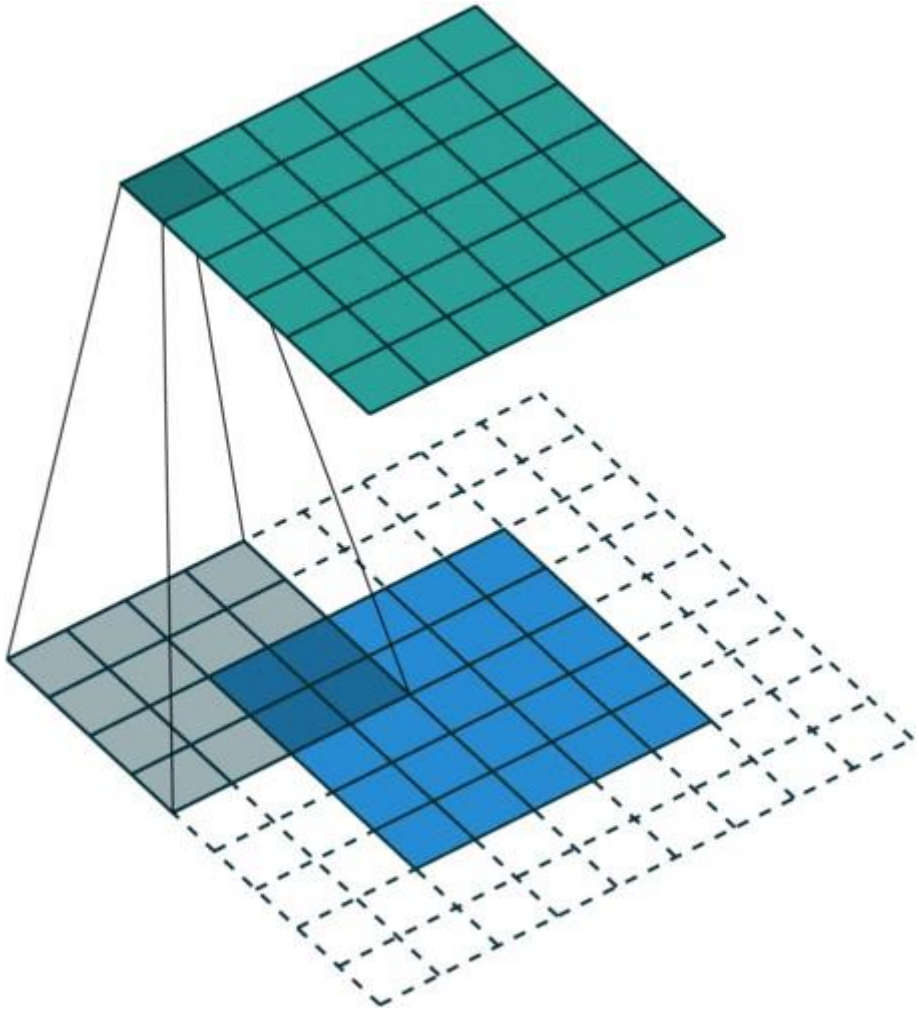


Figura 2. Ejemplo de una convolución. El filtro convolucional tamaño 4x4 (en gris) va recorriendo la imagen (en azul) y generando la imagen filtrada (en verde). Si bien el ejemplo trabaja con un único canal de entrada, podrían ser más. Además se suelen utilizar múltiples filtros por capa, deviniendo en más de 1 imagen (o canal) de salida.

Nótese que se trata de un proceso anidado, en que únicamente la primera capa de filtros (o *kernel)* es aplicada directamente sobre (los canales de) la imagen original, mientras que las capas subsiguientes se aplican sobre la representación obtenida de la misma tras haber sido “filtradas”

por todas las capas anteriores. Estas representaciones se conocen como canales de salida o *feature maps*, y es debido a su naturaleza recursiva que es posible hablar de una jerarquía de representaciones, en donde las primeras capas detectan patrones más específicos, como pueden ser líneas o curvas, mientras que las últimas son más especializadas, detectando formas más complejas como por ejemplo lo que para el ser humano sería un rostro o una silueta.

Las convoluciones son seguidas normalmente por funciones de activación como la *ReLu*, de modo de introducir un componente no lineal en la misma y así ganar en expresividad de la función compleja que se está representando, y también por capas de *Pooling* (*Max* o *Average*, las más comunes) cuyo principal objetivo es reducir el tamaño de las sucesivas imágenes y con ello disminuir la cantidad total de parámetros de la red, que crece exponencialmente a medida que se agregan más y más capas.

Por último, es característico agregar a la salida del último bloque convolucional una o más capas densas o *fully-connected* que rompen la estructura espacial de la imagen, convirtiéndola en un vector unidimensional y permitiendo así concretar el propósito planteado inicialmente, como puede ser clasificar la imagen original como perteneciente a una determinada clase.

El proceso completo queda ilustrado en la figura debajo. La imagen ingresa, es filtrada por sucesivos conjuntos de convoluciones 2D y *poolings,* generando canales de salida a su paso, para finalmente aplanar los de la última capa en un vector de 1 dimensión, que tras pasar por una serie de capas densas resulta en una salida, compuesto por tantas neuronas como clase tiene el problema. Los pesos de los filtros convolucionales serán aprendidos por la red para lograr la mejor *performance* posible, aquí y en forma simplificada que la neurona que identifica a la imagen como perteneciente a una cebra, la segunda de la capa de salida, tenga un valor tan alto como sea posible. La función de activación de la última capa debe corresponder con el tipo de ejercicio, utilizándose normalmente la *sigmoidea* para clasificaciones binarias y la *softmax* para problemas multiclase.

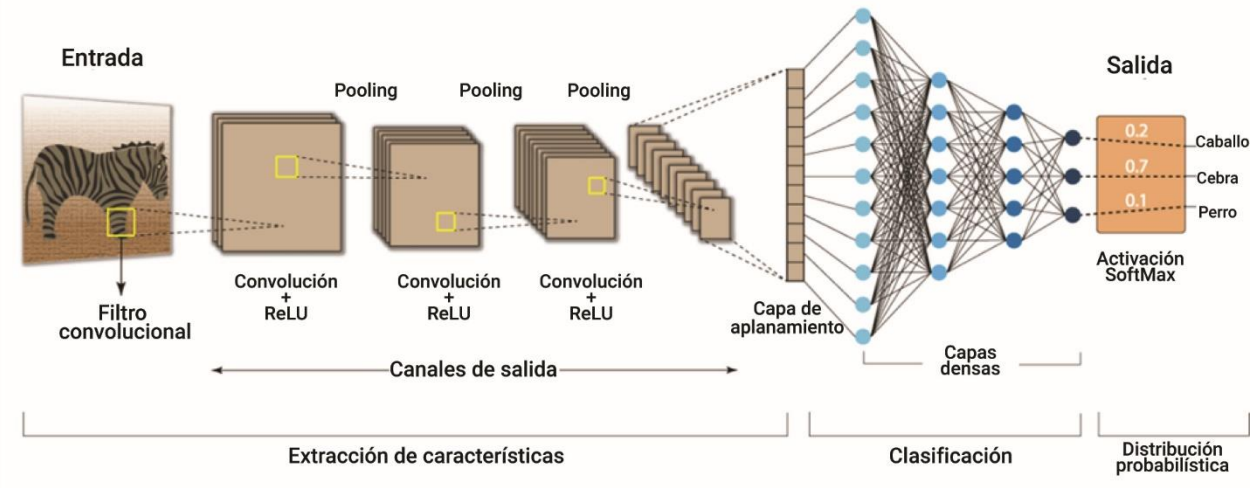


Figura 3. Arquitectura de una CNN para un problema de clasificación. Nótese que únicamente el primer filtro convolucional opera sobre la imagen de entrada, mientras que los subsiguiente operan sobre las representaciones que de ella hicieron los filtros anteriores. Además, que es preciso aplanar la última representación para poder, varias capas densas mediante, ofrecer una predicción.

Por último debe mencionarse que si bien los pesos de los filtros convolucionales aprendidos por la red resultan del entrenamiento para resolver un problema en particular, los patrones que estos identifican resultan a menudo útiles para aprehender otros problemas relacionados, sobre todo cuando se trata de arquitecturas profundas (de varias capas) y entrenadas con un *dataset* de tamaño considerable. Este proceso se conoce como transferencia de aprendizaje o t*ransfer learning*, y consiste en aprovechar los patrones reconocidos por modelos preentrenados, lo que se traduce en congelar los parámetros de varias o todas sus capas convolucionales, agregando únicamente más capas al final del modelo de modo de adaptar el diseño de la red al nuevo problema.

2. Objetivo general y específicos

El presente trabajo se propone desarrollar un clasificador eficiente para detectar las 2 patologías del ojo humano antes presentadas, glaucoma y cataratas, a través de la utilización de Redes Neuronales Convolucionales. Cabe destacar que estas enfermedades no son excluyentes entre sí, por lo que se está frente a un problema multi-etiquetas, contraponiéndose estas a ojos considerados "sanos", es decir a globos oculares que no presentan estas (ni otras) alteraciones de acuerdo al diagnóstico (etiquetas) hecho por especialistas humanos.

Adicionalmente se plantean como objetivos particulares:

- Evaluar si es posible obtener una buena *performance* en la identificación de las enfermedades arriba mencionadas, comparado contra varios casos base o *dummy*.
- Cotejar la *performance* de distintas arquitecturas de *CNNs*, particularmente la de aquellas construidas manualmente con la de redes más profundas preentrenadas con otros conjuntos de datos, a las cuales ocasionalmente se les reentrena su última capa.
- Identificar el mejor modelo y explorar si las partes de la imagen que más contribuyen a la clasificación que este arroja se corresponden con el conocimiento que se tiene *a priori* respecto a qué parte del ojo debería verse afectada en cada enfermedad.
- Precisar si es posible mejorar los resultados incorporando información adicional a la de las imágenes, como lo son aquí la edad y el género del paciente.

3. Conjunto de datos

El conjunto de datos utilizado fue publicado en el sitio web Kaggle en abril de 2020[ii], que a su vez fue tomado de una competencia publicada por la Universidad de Pekín en el año 2019, conocida por sus siglas ODIR (Ocular Disease Intelligent Recognition)[iii].

El *dataset* se compone originalmente de 5,000 pacientes con las respectivas imágenes a color de sus fondos de ojo, pero sólo para 3,500 se conoce su edad, género y las palabras claves de los especialistas, por lo que solo esas son de utilidad aquí. Fue recolectado por la empresa Shanggong Medical Technology Co., Ltd. de diferentes hospitales y centros médicos de China, en los cuales las imágenes se tomaron utilizando una variedad de marcas de cámaras existentes en el mercado, como Canon, Zeiss, Kowa, entre otras, resultando en tamaños y calidades dispares.

Los diagnósticos médicos devinieron en la clasificación de los ojos con 8 etiquetas, que se incluyen en igual número de columnas del *dataset*, a saber:

- Normal (N)
- Diabetes (D)
- Glaucoma (G)
- Catarata (C)
- Degeneración macular (A)
- Hipertensión (H)
- Miopía patológica (M)
- Otras enfermedades (O)

Se destaca que únicamente la primera, referente a un ojo completamente sano, es excluyente respecto al resto; esto es, ambas enfermedades, a priori, pueden darse en forma concomitante en el mismo ojo.

4. Metodología

4.1. Análisis exploratorio

4.1.1. Pacientes

La exploración inicial de muestra que la edad promedio del *dataset* es de 57.8 años, mientras que la mediana es ligeramente superior, de 59 años, su desvío estándar es de 11.7 años y su rango intercuartil de 15 años. La distribución parece bastante simétrica, aunque se observan algunos *outliers* cercanos al 0.

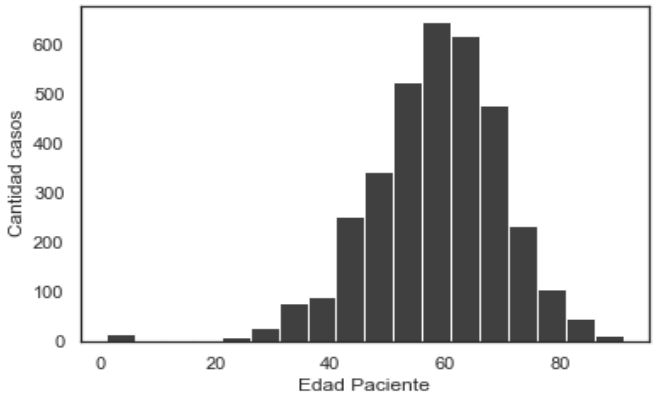


Figura 4. Histograma de la edad de los pacientes bajo estudio. Se observan algunos outliers en la parte izquierda, cercano al 0.

Un análisis en detalle de tal anomalía demuestra que se cuenta con 16 casos, todos de sexo femenino, en que la edad es de tan solo 1 año. Debido a lo atípico de la situación parecen más bien errores de tipeo o alguna codificación desconocida, que casos reales de bebes. En el punto 5.3., en el que se incorpora esta información en el modelo predictivo, se utilizará la media de las mujeres en reemplazo de estos valores atípicos, de modo de remediar la situación.

En términos de género el *dataset* se encuentra balanceado, siendo la proporción de individuos masculinos ligeramente superior (54 a 46%).

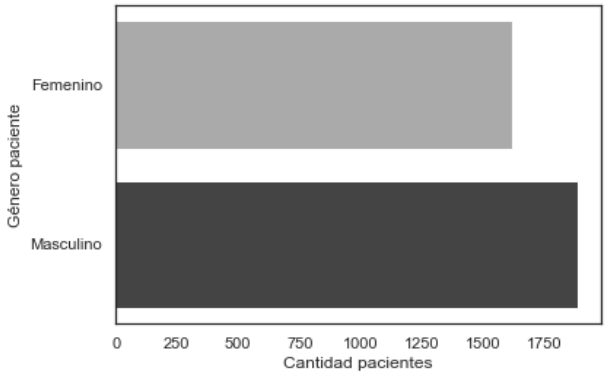


Figura 5. Cantidad de pacientes por género bajo estudio. La cantidad de hombres es ligeramente superior.

La interrelación de ambas variables tampoco arroja diferencias significativas, resultando ambas simétricas en términos de la edad, como se veía en la gráfica general, aunque la media y mediana de las mujeres es levemente superior, aunque pese a tener los *outliers* de 1 año arriba mencionados.

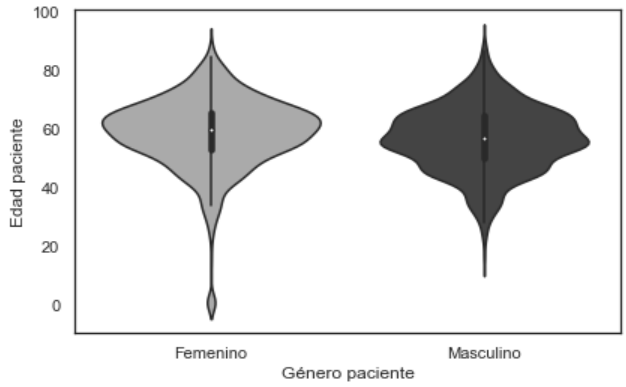


Figura 6. Interrelación entre las variables de género y edad. La forma parecida de las distribuciones dan cuenta que no existen diferencias sustanciales en la edad si se separa por género.

A continuación, se evalúo la relación de estas 2 variables con la condición estar enfermos, lo que se traduce en tener al menos 1 enfermedad en alguno de sus ojos. A simple vista no parece haber una influencia marcada de la edad y género en tal condición, pero ello no evita que pueda serlo para enfermedades particulares o en presencia de ciertos patrones en las imágenes de fondos de ojo, por lo que se dejará al modelo llevar a cabo tal juicio.

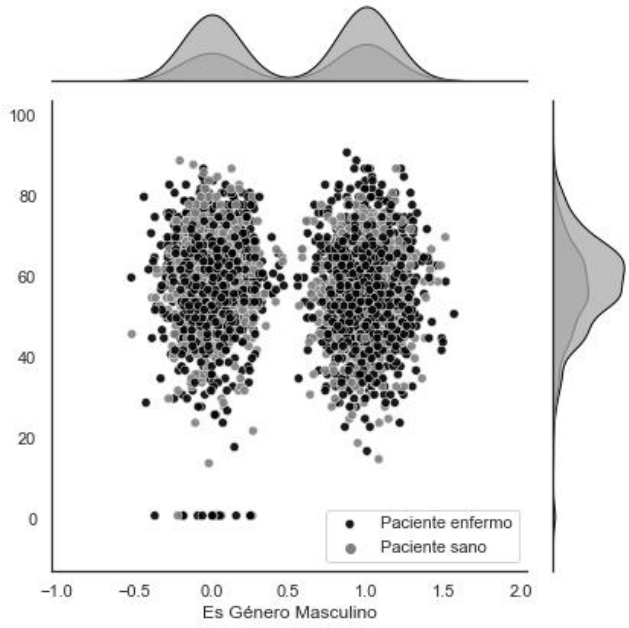


Figura 7. Interrelación de las variables de género (aquí como booleana y con jittering) y edad con la condición de pacientes con ojos enfermos. No se observa un patrón claro que permita suponer que poseen poder explicativo.

En términos de las condiciones de los pacientes, de los 3500 analizados, únicamente 1140, alrededor del 32%, contaba con la etiqueta "N", indicativa de que no poseían enfermedades en ninguno de los dos ojos. El desglose de los 2360 casos restantes se detalla en la Tabla 1.

Enfermedad	Pacientes
Diabetes (D)	1128
Glaucoma (G)	215
Cataratas (C)	212
Degeneración Macular (A)	164
Hipertensión (H)	103
Miopía (M)	174
Otras enfermedades (O)	979
Total	2975

Tabla 1. Cantidad de pacientes con las distintas enfermedades bajo estudio.

Al existir pacientes con más de 1 enfermedad, incierto a priori si en el mismo ojo o repartidas entre ambos, el total de enfermedades contabilizadas, 2975, es ligeramente superior a la cantidad de pacientes enfermos. Si se analiza este número, se observa que más de la mitad de los sujetos bajo análisis padecía una única condición, mientras que poco menos del 16% (557 casos) tenían 2 enfermedades concomitantes y solo en casos contados coexistían 3 patologías (29 casos).

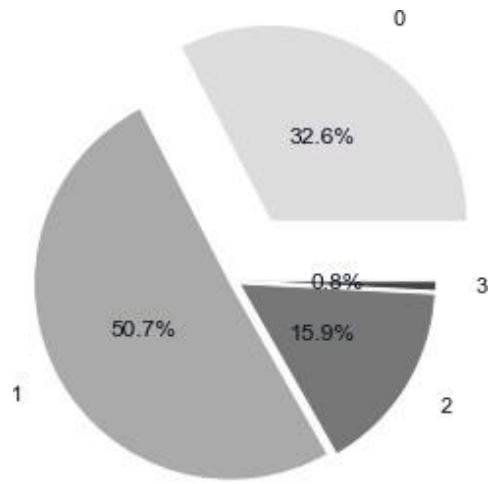


Figura 8. Porcentaje de pacientes por número de enfermedades que padecen. Aproximadamente un tercio tiene ambos ojos sanos mientras que la mitad del total poseen solo 1 condición.

Desagregando dicha información por enfermedad y normalizando por el total marginal de cada una, se distingue que la coexistencia de 1 o 2 enfermedades adicionales a la utilizada para agrupar es común a todas las patologías.

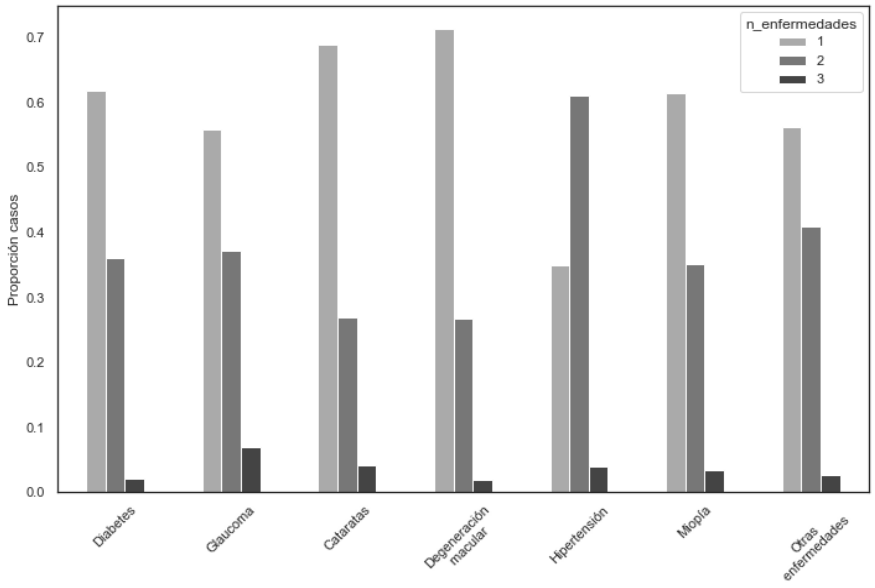


Figura 9. Proporción de pacientes por número de enfermedades, abierto por etiqueta de enfermedad; en naranja cuando la enfermedad del eje X es la única existente. Exceptuando la hipertensión, los patrones son similares.

Esto es marcadamente superior en el caso de la hipertensión, donde el porcentaje del total de casos en que existe como enfermedad única es del 35% mientras que en el 61% de los casos coexiste con 1 enfermedad adicional; en el resto de las enfermedades estos números se ubican, en promedio, en el 63% y 34% de los casos, respectivamente.

El siguiente paso naturalmente resulta el análisis de las coocurrencias de las enfermedades en un mismo paciente. Se recuerda nuevamente que al tratarse hasta aquí del análisis de pacientes y no de ojos por separado, esto puede dar una pauta de correlaciones entre patologías a nivel persona que no necesariamente resultan trasladables al caso de un mismo ojo, siendo esta última un caso particular del primero.

La matriz normalizada resulta asimétrica, en la que cada fila debe leerse como la ocurrencia de una enfermedad y las columnas que la intersecan las enfermedades concomitantes. La intersección, esto es, cada celda, debe leerse como la proporción de casos de la enfermedad de la fila que también poseen la enfermedad de la columna. A modo de ejemplo, existen 32 pares de ojos con diabetes y glaucoma, lo cual representa tan solo un 2.8% del total de ojos con diabetes (1128) pero un 14.9% de los ojos con glaucoma (215).

	N	D	G	C	A	H	M	O
N	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
D	0.0	100.0	2.8	3.5	1.4	4.0	1.7	26.9
G	0.0	14.9	100.0	1.4	5.1	3.7	3.3	22.8
C	0.0	18.4	1.4	100.0	0.0	0.9	0.0	14.6
A	0.0	9.8	6.7	0.0	100.0	2.4	1.8	9.8
H	0.0	43.7	7.8	1.9	3.9	100.0	0.0	11.7
M	0.0	10.9	4.0	0.0	1.7	0.0	100.0	25.3
O	0.0	30.9	5.0	3.2	1.6	1.2	4.5	100.0

Tabla 2. Matriz de coocurrencia de enfermedades (codificadas). Obsérvese que es asimétrica: cada celda marca la proporción del total de pacientes que tienen la enfermedad indicada en las filas que también padecen la enfermedad indicada en las columnas.

Nótese luego que (leyendo la columna “D”) la diabetes aparece marcadamente en conjunto con todas las demás enfermedades, especialmente con la hipertensión (43.7% de los ojos hipertensos presentan diabetes) y a las "otras enfermedades" (30.9% de los ojos con ellas presentan también diabetes). Estás últimas también aparecen asociadas a casi todo el resto, particularmente a la diabetes (el otro lado de la moneda, 26.9% de los ojos con diabetes también presentan "otra enfermedad"), miopía (25.3% de los ojos con miopía presentan "otra enfermedad") y glaucoma (22.8% de los ojos con glaucoma presentan "otra anormalidad").

Para las enfermedades de interés, cataratas y glaucoma, no se observa marcada coocurrencia a nivel paciente, siendo ambas cifras gemelas en torno a 1.4 de cada 100 pacientes; existen únicamente 32 pacientes con ambas enfermedades.

4.1.2. Fondos de ojo

Las 7,000 imágenes a utilizarse, dos por paciente, están todas comprimidas en formato JPEG, tienen 3 canales, RGB, y tal como se anticipó y se ve debajo, tienen diversos tamaños.

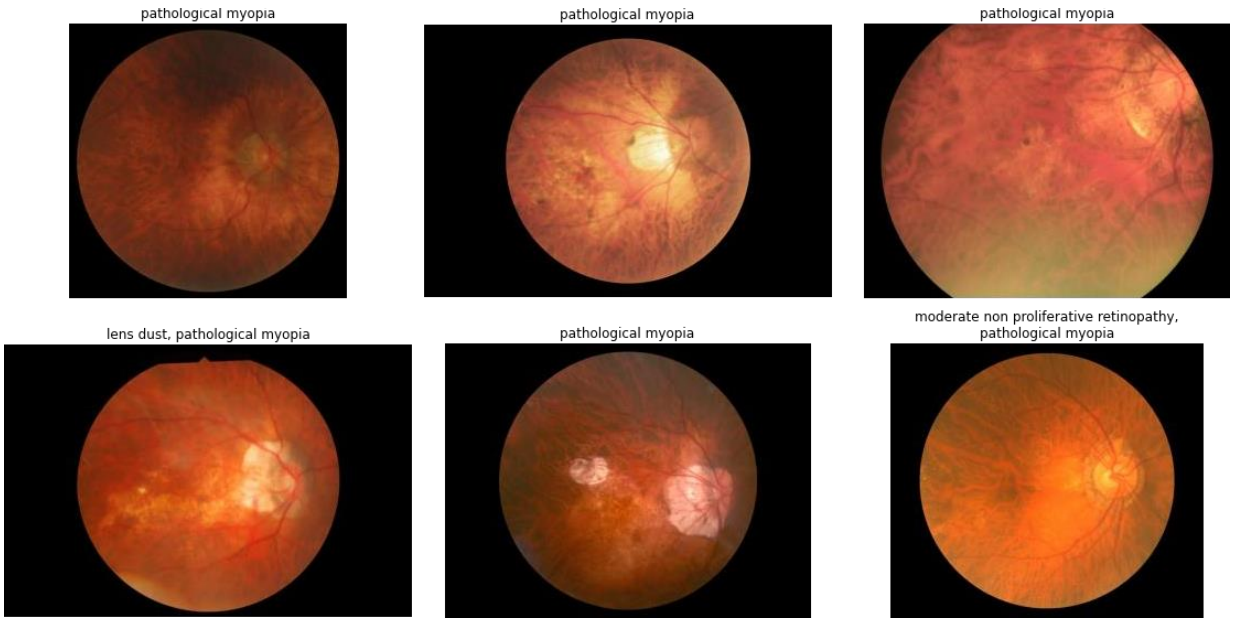


Figura 10. Imágenes de muestra de los fondos de ojo. Las mismas corresponden a 6 ojos que padecen miopía. Nótese la diferencia en los tamaños originales.

En su mayoría (el 88%) tienen forma horizontal (más anchas que altas), seguidos de algunas de formato cuadrado y, más raras, forma vertical. Redondeando a 256 píxeles, la Tabla 3 muestra que predominan imágenes del orden de 2560x1536 pixeles.

ancho	0	256	512	768	1280	1536	1792	2048	2304	2560	2816	3072	3328	3584	3840	4096	4352	5120
alto																		
0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
256	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
512	0	0	5	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
768	0	0	0	13	12	0	94	0	0	0	0	0	0	0	0	0	0	0
1024	0	0	0	0	34	265	14	0	0	0	0	0	0	0	0	0	0	0
1280	0	0	0	0	357	8	369	20	0	0	0	0	0	0	0	0	0	0
1536	0	0	0	0	0	4	0	545	480	2146	0	0	0	0	0	0	0	0
1792	0	0	0	0	0	0	390	21	15	627	292	0	0	0	0	0	0	0
2048	0	0	0	0	0	0	0	187	1	0	0	18	0	0	0	0	0	0
2304	0	0	0	0	0	0	0	0	98	6	0	202	296	30	0	0	0	0
2560	0	0	0	0	0	0	0	0	0	0	0	0	0	0	124	0	0	0
2816	0	0	0	0	0	0	0	0	0	0	194	0	0	0	0	16	32	0
3328	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	52

Tabla 3. Tamaño original de las imágenes bajo estudio, redondeadas a 256 pixeles. La mayoría se concentra alrededor de un ancho de entre 2048 y 2560 y un alto de ente 1536 y 1792.

Se está por tanto a imágenes de calidad alta, que será necesario reducir para poder utilizarlas en los modelos a construir debido a las limitaciones en la capacidad de cómputo del equipo utilizado.

4.2. Preprocesamiento

4.2.1. Separación de etiquetas por ojo

Para hacer más precisa la asociación entre una patología y la imagen correspondiente, y al mismo tiempo duplicar la cantidad de casos disponibles, se toma la decisión de separar las filas por ojo.

La dificultad que ello presenta es que las etiquetas de clase están colocadas por paciente, sin precisar a cuál de los 2 ojos pertenecen. La tarea de colocar etiquetas a nivel ojo usando la información disponible no resulta trivial, debido a que:

- Si se está frente a una fila con una única enfermedad, se presenta el desafío de entender si la misma responde a que ambos ojos están enfermos o si solo uno de ellos.
- El problema se magnifica cuando una fila tiene marcadas 2 enfermedades, puesto esto puede resultar de una enfermedad en cada ojo, las dos solo en un ojo mientras que el otro tiene solo 1 o está totalmente sano, o bien ambos ojos tienes ambas enfermedades.
- Esto se vuelve aún más complejo para el caso de 3 etiquetas, aunque son casos aislados en nuestro *dataset*.

Para poder obtener tal resultado se procedió a evaluar los diagnósticos existentes, identificar aquellos que se asocian inequívocamente con una patología y usar ese diccionario para clasificar a cada ojo con las respectivas patologías que cada uno padece.

Como puntapié inicial, se vislumbra que la estructura de los diagnósticos es siempre similar, con "frases claves" separadas por comas, de haber más de 1 una. Por tanto, parece mejor extraer las mismas para intentar asociarlas a enfermedad(es), que las palabras individuales.

Asimismo solo se considerarán aquellos diagnósticos asociados con una única condición, sea un ojo sano o con una única patología, dado que de lo contrario no sería posible asociar cada diagnostico (y sus respectivos frases claves) a una única condición. Por ejemplo, una frase que aparece en un ojo que tiene tanto miopía como cataratas tendría que ser contabilizada en ambas patologías, sin permitir diferenciar a cuál de ellas se asocia. Esto no resulta un gran problema dado que, como se dijo antes, estos casos representan el 83% del total.

De este ejercicio resultan la siguiente cantidad de frases claves, y su respectiva cantidad de ocurrencias.

Condición	Frases clave	Número de ocurrencias
Normal (N)	3	1277
Diabetes (D)	62	2085
Glaucoma (G)	39	400
Cataratas (C)	20	373

Degeneración Macular (A)	22	259
Hipertensión (H)	23	178
Miopía (M)	27	298
Otras enfermedades (O)	91	2150

Tabla 4. Cantidad de frases claves y número de ocurrencias (ojos) de las mismas por cada enfermedad considerada.

Se cuenta entonces con al menos una frase diagnóstica clave por patología, habiendo muchas más para la categoría de "Otras enfermedades", como era de esperarse, debido a que es una categoría residual.

Quitando frases generales que no parecen asociadas a la condición si no a defectos en la imagen y que incluso aparecen en más de 1 clase (como "lens dust" o "low image quality"), se hallan patrones marcados en las distintas categorías, a saber:

- Normal: el 99.9% de los casos quedan cubiertos por la frase "normal fundus".
- Glaucoma: sumando los 2 primeros términos, "glaucoma" y "suspected glaucoma", indiscutiblemente típicos de esta condición, se alcanza un 96% de los ojos considerados como tales, de no haber solapamiento entre frases.
- Catarata: más del 96% de las imágenes con esta condición tienen el diagnóstico de "cataract".
- En el resto de las categorías se alcanzan porcentajes igualmente altos.

Más importante aún resulta que:

- Estas frases seleccionadas no aparecen asociadas a ninguna otra enfermedad y por tanto su "poder predictivo" es total.
- De los casos con más de 1 diagnóstico, un total de 1172 ojos (o 586 pacientes) no utilizadas para construir esta asociación, casi la totalidad (1157, el 98.7%) presenta alguna de estas frases identificadas y por tanto será posible asignarles etiquetas.

Con este diccionario definido, se procedió entonces a la clasificación de cada ojo usando sus correspondientes frases diagnósticas. Como parte del mismo se debieron remover 36 ojos debido a que no tenían ninguna etiqueta asignada, y remover la etiqueta (excluyente) "N" de 4 ojos que también fueron clasificados como "O" por contener también una frase característica de este último grupo. El resultado es un dataset de 6964 filas, 3483 ojos izquierdos y 3481 ojos derechos, con la siguiente cantidad de casos:

Condición	Ojos	Comp. Con Pacientes
Normal (N)	3095	2.72
Diabetes (D)	1801	1.60
Otras enfermedades (O)	1188	1.21
Glaucoma (G)	326	1.52
Cataratas (C)	313	1.48
Degeneración macular (A)	280	1.71
Miopía (M)	268	1.54

Hipertensión (H)	193	1.87
Total	7464	2.51

Tabla 5. Cantidad de ojos por enfermedad, una vez termina el proceso de separación de etiquetas por ojo.

Como se puede ver en la segunda columna de la Tabla 5, la relación con la cantidad de ojos del *dataset* inicial se mantiene por debajo de 2, dado que ahora no se contabilizan para cada condición ambos ojos si no solo aquellos que presentan diagnósticos asociados. La excepción son los ojos sanos, dado que antes la etiqueta no se colocaba a menos que ambos ojos estuvieran sanos.

Para el presente trabajo, centrado en el análisis de cataratas y/o glaucoma, el *dataset* final queda compuesto por 3732 ojos: 3095 sin patologías, 324 solo con glaucoma, 311 solo con cataratas y 2 casos de enfermedades concomitantes.

4.2.2. Preprocesamiento de imágenes

El preprocesamiento de la imagen consistió en el recorte de los bordes de la imagen, normalizar la imagen de modo incrementar su contraste y así eliminar algo de ruido, y posteriormente modificar su tamaño a una forma cuadrada, de 256 x 256 pixeles.

4.3. Clasificador

4.3.1. Entrenamiento y selección de *CNNs*, y comparativa con escenarios base

El siguiente paso fue el entrenamiento de los modelos de *CNNs*, la comparación entre sí y con otros casos bases, y la elección del que mejor performance ofrecía, utilizando alguna medida particular predeterminada.

Se debe resaltar que se está frente a un problema multietiqueta aunque bastante restringido, debidoa que:

- Como recién se marcó, si bien las cataratas (“C”) y el glaucoma (“G”) pueden darse al mismo tiempo en el mismo ojo, existen muy pocos ejemplos en que esto es así.
- La restante categoría, ojo sano (“N”) sí resulta excluyente respecto a las otras 2, y puede marcarse aquí simplemente como la ausencia de etiquetas en estas últimas.

En primera instancia se separó un 20% del *dataset* original para ser utilizado como conjunto de Test. A continuación, con el 80% restante, se procedió al tándem entrenamiento-validación utilizando una estrategia de *5-Fold Cross Validation*, promediándose la performance en los distintos *folds* de validación para arribar a una decisión final. Sobre este último es preciso mencionar que:

- La separación en *Folds* se hizo en forma estratificada, de modo de mantener, en la medida de lo posible, la proporción de clases dentro de cada uno.
- Se entrenaron 5 arquitecturas distintas, listadas a continuación:
    - Un diseño propio (denominado *OwnModel*), constituido por sucesivas 4 capas de convoluciones 2D + Max Pooling, y coronado por un *Flatten* más 2 capas densas, con activación ReLu, previo a la capa de salida. Este cuenta con aprox. 784 mil pesos entrenables.
    - El modelo VGG19[iv], de gran profundidad, preentrenado con el *dataset imagenet* y su última capa removida; su salida de 512 imágenes (canales) tamaño 8x8 también se aplana, y se la pasa por una capa densa, también con activación ReLu, antes de dar la predicción. De este se cuenta con dos variantes:
        - Todas las capas VGG congeladas, siendo la capa densa agregada la única entrenable (denominado *VGG+dense64*). Cuenta con aproximadamente 2 millones de pesos entrenables.
        - La última capa de Convolución 2D, conocida como *"block5_conv4"*, también entrenable, dándole así a la red mayor flexibilidad a partir de más parámetros factibles de ser ajustados (denominado *VGG+dense64+trainableLastLayer*). Posee por tanto alrededor de 4.4 millones de pesos entrenables.
    - Mismo par con el modelo preentrenado ResNet50[v], que hace uso del concepto de bloques residuales, con pesos de *ImageNet*. Esto es, uno con y uno sin la posibilidad de entrenar la última de sus capas (aquí el par Convolución 2D + Max Pooling denominado "*conv5_block3_3*"), para luego, otra vez, aplanar los 2048 *feature maps* tamaño 8x8 resultantes, agregar una capa densa con activación ReLu y obtener las predicciones (denominados *ResNet50+dense64+trainableLastLayer* y *ResNet50+dense64*, con aproximadamente 8.4 y 9.4 millones de pesos entrenables cada una, respectivamente).

  En todos los casos, la capa de salida cuenta con 2 neuronas, una por enfermedad, con activación sigmoidea, dado que los resultados no son mutuamente excluyentes. El diseño de cada una puede verse *notebook* anexo.

- La función de pérdida escogida fue la Entropía Cruzada Binaria, debido a que cada predicción hecha, 2 por cada muestra, era binaria.
- Para cada una se utilizó el optimizador *Adam*, con dos *learning rates* distintos, 1E-5 y 1E-4.
- Debido al desbalance de clases existente, contándose con muchos más casos de ojos sanos que de ojos con cualquiera de las otras patologías (83% del total resulta clase negativa, frente a un 9% de glaucoma y 8% de cataratas), se decidió utilizar la estrategia de darle peso a los distintos ejemplos, de modo de penalizar en forma inversamente proporcional a la proporción de clase, al error de la Red Convolucional.
- También por esta sobrepoblación de ojos sin patología alguna se decidió utilizar el *Macro Avg. F1-Score* como métrica de performance para obtener el mejor modelo. Esta métrica

hace un promedio simple de los F1-Score de cada clase (aquí 2, una por cada enfermedad) calculados en forma independiente, y por tanto resulta útil para asegurarnos que la red no cometa demasiados errores, sean falsos positivos o negativos, en la identificación de ninguna de las 2 enfermedades.

- En cada uno de los 50 procesos de entrenamiento (5 Folds, 5 modelos y 2 tasas de aprendizaje) se hizo uso de un *batch size* de 64 imágenes durante un máximo de 20 épocas. Máximo refiere aquí a que podían ser menos ya que se estableció un mecanismo de *Early Stopping* sobre la pérdida del conjunto de validación; esto es, si la misma no disminuía durante las siguientes 2 épocas el entrenamiento se detenía.

El resultado promedio de cada experimento fue comparado con clasificadores *Dummy,* de modo de contar con un caso base, a saber:

- Caso más frecuente: todos los ojos están sanos (denominado *allHealthy*). Debido a que la clase doblemente negativa es la predominante, esto sería estadísticamente lo más probable, sin embargo no tiene éxito alguno en el propósito original de detectar enfermedades.
- Caso más infrecuente (denominado *allUnhealthy*): todos los ojos tienen ambas enfermedades. Esto es altamente improbable, de acuerdo a las proporciones originales, pero sirve para poner, blanco sobre negro, los resultados obtenidos.
- Caso proporcional estratificado (denominado *Proportional*): se asignan al azar, cada clase por separado, la proporción de etiquetas con que se cuenta originalmente en el *dataset*. Es un escenario menos extremo.

Una vez obtenido el mejor modelo, de acuerdo al criterio arriba especificado, se procede a reentrenar la arquitectura con todos los datos de entrenamiento, durante una cantidad de épocas alineadas con los resultados del *Early Stopping* y se predicen las etiquetas del conjunto de *test* y miden sus resultados.

4.3.2. Zonas relevantes para la clasificación

A continuación, a modo exploratorio, se intenta entender, para cada una de las clases, qué zonas de la imagen influencian más al modelo para tomar una decisión correcta, con el afán de encontrar algún parangón con los patrones característicos de las enfermedades bajo estudio.

Para ello se hace uso de una técnica conocida como *Gradient-weighted Class Activation Mapping* (*Grad-CAM*) que, una vez completo el entrenamiento de la red y fijados los parámetros, utiliza los gradientes del *output* escogido, que fluye hacia la última capa convolucional para producir un mapa de localización que resalta las regiones importantes de la imagen a la hora de predecir el resultado. En otras palabras, permite visualizar dónde está "mirando" la *CNN* al arrojar una predicción.

En este caso, se tomarán, por cada clase, todos los ejemplos positivos clasificados correctamente y se hará un promedio ponderado, en base a la probabilidad de la predicción, de todos los mapas. Cabe remarcar que esto aquí cobra sentido porque las imágenes están tomadas desde un ángulo similar, pero de no ser así sería más útil analizar cada caso en forma aislada.

4.3.3. Utilización de datos adicionales

Como último experimento, se agrega una segunda entrada a la arquitectura con mejor *performance*, la edad y el género del paciente cuyo ojo está siendo analizado.

En ese afán primero se debió sobrescribir con la media del género correspondiente a los valores anómalos detectados en el análisis exploratorio, dado que parecía raro por las características de la muestra que se tuviera muestras de bebes de 1 año de edad. Luego se normalizó el rango etario a 0-1, para que coincidiera con las demás *features*.

Respecto al género, fue necesario codificarlo como variable binaria, "Is Male", indicando que se trataba (o no) de un ojo perteneciente a un masculino.

A la capa de entrada de 2 *features* se la pasó por una capa densa de tamaño 8, para darle expresividad a la red en la utilización de estos nuevos datos, y luego se concatenó estas 8 salidas a las 64 resultantes de la anteúltima capa densa de la rama convolucional del modelo.

Con este diseño se corrió un proceso de validación cruzada con 5 *folds*, similar al anterior, variando la tasa de aprendizaje, para detectar el mejor ejemplar y luego obtener la *performance* contra el *dataset* de *test* y compararla contra el modelo de mejor resultado pero sin esta información de los pacientes agregada.

5. Resultados y discusión

5.1. Selección del mejor modelo y performance en *dataset* de *test*

A continuación se presentan los resultados de los modelos del punto 4.3.1., ordenados de mejor o peor *performance* del Macro F1-Score promedio del *Fold* de validación.

Modelo	Tasa de aprendizaje (*Adam*)	Validación – Macro F1 Score promedio	Entrenamiento – Macro F1 Score promedio	Promedio de épocas corridas
ResNet50+dense64	1E-4	0.632	0.965	10.4
ResNet50+dense64+trainableLastLayer	1E-4	0.614	0.940	6
ResNet50+dense64	1E-5	0.606	0.961	8.2
VGG+dense64	1E-4	0.591	0.953	13.6
vgg+dense64+trainablelastlayer	1E-4	0.581	0.928	5.8
resnet50+dense64+trainablelastlayer	1E-5	0.580	0.899	6.6
vgg+dense64	1E-5	0.544	0.946	13.8
VGG+dense64+trainableLastLayer	1E-5	0.503	0.965	7.4
OwnModel	1E-4	0.224	0.239	7
OwnModel	1E-5	0.199	0.204	10.2
AllUnhealthy	NA	NA	0.158	NA
Proportional	NA	NA	0.087	NA
AllHealthy	NA	NA	0	NA

Tabla 5. Comparativa de performance, en validación y entrenamiento, de los distintos modelos, con sus respectivas tasas de aprendizaje.

Como se observa, el mejor modelo resulta la *CNN* preentrenada *ResNet50*, con todas sus capas congeladas y una capa final densa de 64 neuronas antes de la predicción final. Debe decirse que la diferencia respecto al segundo y tercer lugar, esto es la misma red pero con diferente tasa de aprendizaje o su última capa descongelada, resulta pequeña, por lo que la aleatoriedad inherente a la inicialización de los pesos permite hablar de un "empate técnico". Sin embargo, el hecho de lograr similares resultados con una capa menos que entrenar da cuenta de que la complejidad de la red sin descongelar esos pesos ya resulta suficiente para aprehender el problema bajo estudio.

La *performance* de todos ellos sobre el *dataset* de entrenamiento se ubicó, en todos los casos, considerablemente por encima de los resultados de validación, siendo casi perfecta, lo que da cuenta de la existencia de *sobreajuste*, pese a haberse activado el mecanismo arriba mencionado de *Early Stopping* configurado en forma estricta, ya que le otorgaba al modelo solo 2 épocas de tolerancia para reducir el error de validación antes de detener el proceso, como sucedió sin

excepción. Esto probablemente se deba al tamaño reducido del *dataset*. Observando con cuidado se nota también que esta parada temprana se da antes en modelos con mayor número de pesos ajustables, ergo mayor complejidad y tendencia al *overfitting*.

Por su parte, los modelos basados en la red *VGG19* quedaron mayoritariamente en mitad de la tabla, con *performances* ligeramente inferiores a las de la arquitectura "ganadora".

Ya ostensiblemente por debajo quedaron los 2 modelos construidos en forma manual, tanto en *training* como en validación, posicionándose apenas por encima del modelo *Dummy* en el que todos los ojos se consideraban doblemente enfermos. El aventurar que se está frente a un caso de *underfitting* o subajuste queda opacado por el hecho de que el proceso de entrenamiento se detuvo, en todos salvo en 1 *fold*, antes de las 20 épocas.

Analizando luego en detalle la performance del mejor modelo en cada una de las 2 clases, se observa un patrón que es común a todos los modelos: la *performance* de la clase Cataratas, aquí un *F1-score* de 0.791, es ampliamente superior a la de la clase Glaucoma, aquí 0.473, indicando que el modelo tiene mucha más dificultad en detectar correctamente esta último. Por el contrario, el modelo *ResNet50+dense64* acierta en la mayor parte de las predicciones de Cataratas.

Si se desglosa esta pobre *performance* en su capacidad de generalización del Glaucoma, el *Recall* promedio correspondiente es de 0.449 y la *precision* asociada es de 0.499, no tan distintas. Esto da cuenta que el modelo comete errores de ambos tipos, con una ligera predominancia de Falsos Negativos sobre Falsos Positivos.

Pese a ello, la capacidad de acertar los casos predichos como enfermos para esta enfermedad, la *precision*, resulta 6 veces mayor que el mejor caso *Dummy*, *AllUnhealthy,* y 10 veces en el caso de las cataratas. En el otro frente, la capacidad efectiva de detectar ojos enfermos no resiste análisis comparativo alguno, dado que resulta del 100% en el caso base de no considerar a ningún ojo libre de enfermedad; si la comparativa en *Recall* se efectúa en cambio contra el modelo *Proportional* se puede hablar de mejoras de similar magnitud: 4.5 y 10 veces superior respectivamente frente al caso base.

El entrenamiento del modelo final, *ResNet50+dense64* con tasa de aprendizaje 1E-4, utilizando ahora todo el conjunto de Entrenamiento fue corrido durante 8 épocas, esto es, 2 épocas menos que el promedio contabilizado durante el proceso de validación cruzada, dado que esa era la "paciencia" del mecanismo de corte temprano allí establecido. Si bien no necesariamente este es la cantidad óptima para el *dataset* extendido, máxime cuando se está hablando de un promedio, es una heurística de utilidad.

Los resultados contra el dataset de test o holdout resultan alentadores. Como muestra la Tabla 6, el valor del Macro F1-Score, 0.664, fue de algunos puntos más que el promedio de validación, siendo de 0.470 para la clase Glaucoma y de 0.860 para Cataratas, alineado también con el desglose por clase en validación presentado antes. El recall macro fue de 0.657 y la precision de

0.623, demostrando que se cometen errores en ambos sentidos, con una ligera inclinación ahora hacia los falsos positivos.

Macro F1-Score	Macro Recall	Macro Precision	Macro F1-Score Glaucoma	Macro F1-Score Cataratas
0.664	0.657	0.684	0.470	0.860

Tabla 6. *Resultado del mejor modelo evaluado contra el dataset de test.*

Los errores cometidos por el modelo quedan más claros aun presentando las matrices de confusión correspondientes. Nótese que se construye una por cada clase, en donde los negativos son con respecto a cada una de las enfermedades.

Glaucoma	Predicho Negativo	Predicho Positivo
Es Negativo	636	46
Es Positivo	31	34

Cataratas	Predicho Negativo	Predicho Positivo
Es Negativo	682	3
Es Positivo	13	49

Tabla 7. Matrices de confusión resultante del mejor modelado contra el dataset de test para cada una de las enfermedades bajo estudio. Obsérvese que únicamente se puede sumar los aciertos de la clase positiva, fruto de no haber ojos en Test con ambas enfermedades al mismo tiempo.

Como es sabido, en el conjunto de Test no existía ninguno de los pocos casos concomitantes, esto es, ojos con ambas enfermedades, por lo que es posible diseñar una matriz conjunta, como si se tratase de un problema de múltiples clases y no múltiples etiquetas; de esta forma, se estaría convirtiendo el problema en detectar o no si el ojo posee alguna enfermedad. Lo relevante es ilustrar que, aunque por los motivos antes expuestos no se haya utilizado como métrica relevante, la *accuracy* del mismo es de 0.881, 5 puntos porcentuales más alta que la resultante de calificar a todo como negativo, la clase mayoritaria, con proporción de 0.830 en el *dataset* de *test*. Esto es debido a que si bien se erró en la predicción de 47 ojos sanos esto queda más que compensado por el acierto en la detección de 85 ojos enfermos.

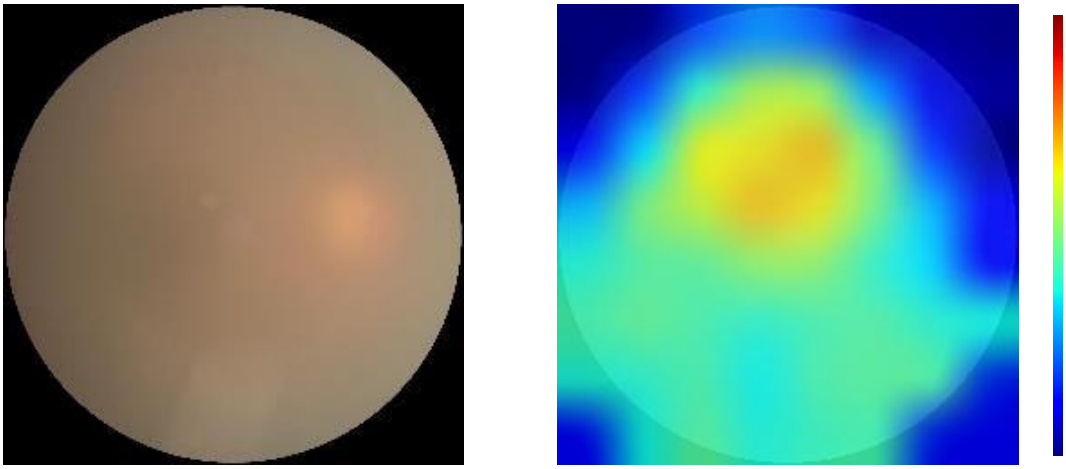
Ojo enfermo	Predicho Negativo	Predicho Positivo
Es Negativo	573	47
Es Positivo	42	85

Tabla 8. Matrices de confusión conjunta, fruto de hacer binario el problema: ojo enfermo o sano.

5.2. Grad-CAM

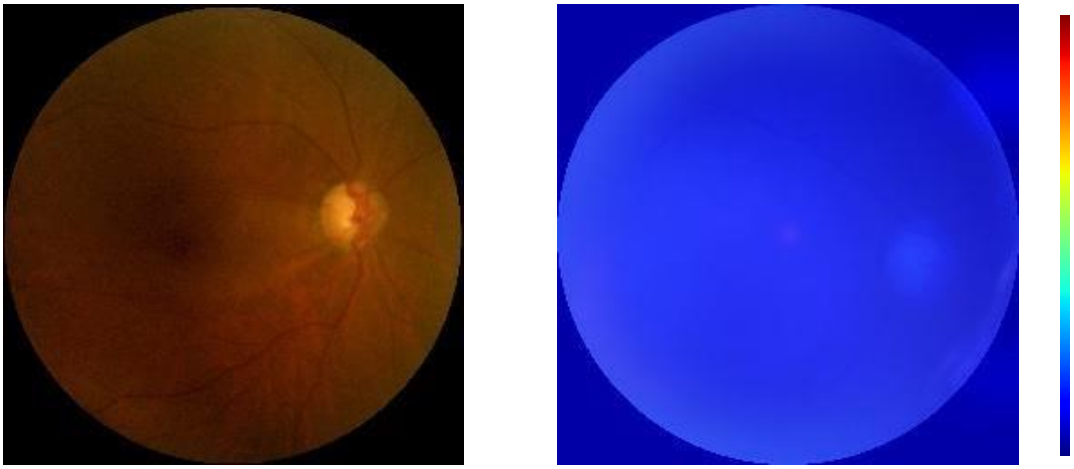
El resultado de los mapas de calor ponderados, correspondientes a los ejemplos positivos de cada clase que fueron clasificados correctamente por el modelo, no muestra similitud alguna con los patrones característicos que estas enfermedades suelen presentar en los estudios de fondo de ojo.

En el caso de las cataratas, como se ve en la figura izquierda debajo, la imagen más característica es una foto de parcial a completamente borrosa, en que no es posible distinguir las diferentes partes del ojo. En el mapa de calor promedio, en la imagen de lado derecho, superpuesto sobre una imagen de un globo ocular para mejor referencia, se ve que la red asigna mayor peso en promedio a la parte central superior de la imagen, marcada en amarillo y a la inferior en menor medida, señalada en tonos verdes (aunque lejos del máximo, que se colorearía de rojo). Esta zona no tiene que ver particularmente con la enfermedad en cuestión, lo cual no quita validez a los resultados obtenidos, dado que, como ya se analizó, lejos están de los guarismos del azar.



Figuras 11a y 11b. 11a. Imagen típica de un fondo de ojo que padece cataratas, con una niebla que impide identificar las partes del ojo. 11b. Mapa de calor promediado de todos los ejemplos positivos bien clasificados por la red. Allí, la parte central del ojo es la que más contribuye a la predicción, algo que no tiene correspondencia con la forma en que los expertos detectan la enfermedad.

Por el lado del glaucoma, el resultado resulta más bien enigmático. El observador especializado detecta el glaucoma mediante cambios en la morfología del nervio óptico, que es el contorno del círculo amarillo que se ve a media altura del lado izquierdo de la Figura 12a. Sin embargo, el mapa de calor promedio del modelo, visible en 12b., es más bien un patrón uniforme de no activación, en el que ninguna parte tiene peso relevante a la hora de clasificar y en donde los bordes del ojo y el nervio óptico se diferencian únicamente por la presencia de la imagen superpuesta de referencia. Dada la performance notablemente inferior en esta clase, es posible que no haya un patrón asociado característico y que más bien la decisión correcta se base en variaciones muy ligeras no perceptibles aquí con la escala utilizada.



Figuras 12a y 12b. 12a. Imagen ilustrativa de un ojo con glaucoma, donde la morfología del nervio óptico (círculo amarillo) se encuentra alterada. 12b. Mapa de calor correspondiente a esta clase. Como se ve, no hay patrón de activación alguno, siendo la forma del ojo y el nervio óptico visibles solo para la contraposición de una imagen de ojo de fondo, para mejor referencia.

5.3. Utilización de datos adicionales

El mejor modelo con el agregado de la información de edad y género de los pacientes se denominó *resnet50+dense64+age+gender* y obtuvo, en su versión con tasa de aprendizaje de 1E-4 y entrenado en promedio durante 7.2 épocas, la mejor *performance* de todos, utilizando como referencia el promedio del *Macro F1-Score* en el *fold* de validación. La diferencia contra *resnet50+dense64* resulta de todos modos pequeña, 0.639 contra 0.632 del original.

Sin embargo esta misma métrica, reentrando el modelo con todos los datos durante 6 épocas y evaluando contra *test,* resultó inferior al original, 0.643 contra 0.664 del modelo sin datos adicionales. Por tanto, no se puede establecer que la introducción de datos adicionales haya producido una mejora en la capacidad de generalización del modelo.

Macro F1-Score	Macro Recall	Macro Precision	Macro F1-Score Glaucoma	Macro F1-Score Cataratas
0.643	0.635	0.655	0.446	0.841

Tabla 9. *Resultado del mejor modelo anterior agregando información de edad y género del paciente, evaluado contra el dataset de test.*

6. Conclusiones y trabajos futuros

En el presente trabajo se ha logrado construir una *CNN* basada en el modelo preentrenado *ResNet50* que permite clasificar, con una *performance* razonablemente buena, las dos patologías más comunes que padece el ojo humano: las cataratas y el glaucoma. Esta afirmación se basa en que, si bien lejos está de la perfección de las etiquetas anotadas por especialistas humanos, las métricas analizadas, particularmente el *F1-Score* promedio, resulta sustancialmente mejor que cualquier de los escenarios base.

De la confrontación de resultados con varias arquitecturas, queda de manifiesto el poderío de utilizar técnicas de *Transfer Learning,* mediante la identificación de patrones preentrenados (disponibles matemáticamente en sus parámetros) en problemas y con datos distintos a los que originalmente motivó su creación y entrenamiento.

No deja de sorprender que aquel modelo que mejor desempeño tuvo fue uno de los que mantuvieron los pesos del modelo preentrenado intactos, resultando en filtros convolucionales y por tanto detección de patrones sin ajuste alguno. Con tan solo añadir 2 capas densas, 1 para dar mayor flexibilidad y otra de salida, se logró un resultado satisfactorio, y ampliamente superior al de la red convolucional creada en forma manual. Más aún, el permitirle reajustar los pesos de su último filtro convolucional, agregándole aún más flexibilidad, no torció la historia, quedando prácticamente con resultados similares en validación y mejorías leves en entrenamiento. Una posible explicación es que la complejidad de la red ya resultaba suficiente para extraer todo el conocimiento que los datos guardaban y por tanto liberar más parámetros ya no tuvo un efecto incremental positivo.

No debe dejarse de mencionar, de todos modos, que la cantidad de datos con la que se trabajó fue relativamente pequeña, sobre todo considerando el gran desbalanceo en favor de la clase negativa, de ojo sano, en contraste con cada una de las clases positivos o enfermedades. Una línea de trabajo futura supone bien conseguir más datos bien utilizar técnicas de *Data Augmentation*, esto es, crear artificialmente más muestras introduciendo transformaciones sobre las imágenes originales.

Asimismo, el haber entrenado una cantidad importante de modelos con recursos computacionales limitados, imposibilitaron la puesta en práctica de experimentos adicionales, a saber:

- Continuar descongelando capas de los modelos preentrenados, sobre todo del modelo ganador, de modo de validar si la *performance* sigue estancada.
- Probar más arquitecturas, tasas de aprendizaje y optimizadores. Esto parece particularmente importante para evaluar las arquitecturas manuales, dado que quizá la creada resultó inapropiada para la complejidad del problema tratado o bien su entrenamiento no fue el óptimo.
- Intentar crear modelos binarios separados por enfermedad, de modo de contrastar los resultados y observar si al hacerlo mejora la *performance*, sobre todo la de la clase Glaucoma, que fue marcadamente pobre. La idea subyacente a esto es que el hecho de

tener que compartir parámetros de la red para identificar 2 enfermedades distintas, casi sin coocurrencias, puede dificultar cuán bien se logra el cometido.

En cuanto a los mapas de calor de las imágenes a la hora de arrojar sus predicciones correctas no se correspondieron en absoluto con el conocimiento humano al respecto. Esto es particularmente interesante en el caso de las cataratas, donde el patrón de activación resulta visible, en tanto que en el caso del glaucoma no se halla estructura alguna, y debería revisarse detenidamente los resultados para ver si es posible encontrar información útil que a primera vista no se vislumbra. Debe decirse al respecto que dicha incongruencia no invalida en absoluto los aciertos y el buen desempeño obtenido, pues solo se pone en juego para intentar darle más transparencia y hacer más explicable a un modelo sumamente complejo, conocido como de *black box.*

Complementando este esfuerzo quedan pendientes algunas tareas relacionadas, como la generación de la imagen ficticia que maximiza el *output* de cada una de las clases, o un análisis en profundidad de los errores de la red.

Por último, el agregado de datos adicionales a las imágenes, como la edad y género del paciente, no lograron mejoraron los resultados previos. Resulta evidente entonces que, como se presagiaba en el análisis exploratorio, estas variables no aportan información adicional a la ya contenida dentro de los fondos de ojo, pero el experimento no deja de resultar interesante, poniendo de manifiesto la posibilidad de combinar información de diferente tipo en una misma *CNN*.

7. Hardware y software

Para llevar adelante el presente trabajo se utilizó una notebook Intel NUC con procesador i5-8259U, 16 GB de memoria RAM y sistema operativo Windows 10 Pro (v. 21H1).

El análisis se corrió íntegramente usando Python 3 como lenguaje, con Jupyter Lab[vi] como IDE y las siguientes librerías:

- *Numpy*[vii] y *Pandas*[viii] para el análisis y manipulación de datos.
- *Matplotlib*[ix] y *Seaborn*[x] para crear visualizaciones.
- *OpenCV*[xi] para el tratamiento de imágenes.
- *Tensorflow*[xii] y *Keras*[xiii] para la construcción y entrenamiento de *CNNs*.
- *Skicit-learn*[xiv] para Splits de datos, construcción de modelos *Dummy* y evaluación de *performance*.
- OS, random, collections, *openpyx*[xv] como ayuda para tareas puntuales.

Anexo

El análisis realizado puede encontrarse en el Notebook "Analisis ODIR" que se adjunta. Debido a la naturaleza estocástica de la inicialización de los pesos de las *CNN* los resultados pueden variar ligeramente.

[i] https://es.wikipedia.org/wiki/Propagaci%C3%B3n_hacia_atr%C3%A1s

[ii] https://www.kaggle.com/andrewmvd/ocular-disease-recognition-odir5k

[iii] https://odir2019.grand-challenge.org/

[iv] https://keras.io/api/applications/vgg/

[v] https://keras.io/api/applications/resnet/

[vi] https://jupyter.org/

[vii] https://numpy.org/

[viii] https://pandas.pydata.org/

[ix] https://matplotlib.org/

[x] https://seaborn.pydata.org/

[xi] https://docs.opencv.org/4.5.2/d6/d00/tutorial_py_root.html

[xii] https://www.tensorflow.org/

[xiii] https://keras.io/

[xiv] https://scikit-learn.org/

[xv] https://openpyxl.readthedocs.io/