Detección de Glaucoma Aplicando Inteligencia Artificial

Rodrigo A. Torres S.
Escuela Profesional Ciencia de la Computación
Universidad Católica San Pablo
Arequipa, Perú
rodrigo.torres.sotomayor@ucsp.edu.pe

Dr. Juan C. Gutiérrez C.
Escuela Profesional Ciencia de la Computación
Universidad Católica San Pablo
Arequipa, Perú
jcgutierrezc@ucsp.edu.pe

***Resumen*—El problema presentado en este artículo fue descubrir la manera más efectiva para la detección del glaucoma a través de la clasificación de imágenes teniendo como herramienta principal la Inteligencia Artificial. Para llevar a cabo esta investigación se analizó distintos trabajos relacionados con el problema mencionado. Se encontró las técnicas Xception, la cual se aplicó a imágenes recortadas automáticamente para así enfocarse en el disco óptico, y el VGG19, donde se utilizaron retinografías para enfocarse en el fondo del ojo. Con este artículo se analizó ambas técnicas utilizadas en el procesamiento de imágenes y se replicarán las pruebas hechas en sus respectivos trabajos con el fin de ser aplicados con una mejora para un trabajo posterior y con los resultados brindar un aporte a la medicina.**

***Palabras Clave* –** Glaucoma, Suport Vector Machine, Inteligencia Artificial, CNN, Xception, VGG19

1. Introducción

En la época actual, uno de los recursos más impresionantes y prometedores para realizar predicciones tanto a corto como a largo plazo es el uso de la Inteligencia Artificial (IA). Esta innovadora tecnología aprovecha técnicas avanzadas de análisis estadístico y matemático para identificar patrones y relaciones en conjuntos masivos de datos, lo que permite anticipar eventos futuros con una gran precisión.

Este artículo se centra en un desafío crítico relacionado con la salud visual que afecta a un número asombroso de personas en todo el mundo. Se estima que más de 2200 millones de individuos sufren de deterioro visual, ya sea en la visión cercana o lejana. Lo que resulta aún más alarmante es que al menos mil millones de estos casos podrían haberse evitado o todavía no se han tratado adecuadamente. Las causas principales de discapacidad visual y ceguera a nivel global son los errores de refracción, glaucoma y las cataratas, condiciones que, en muchos casos, podrían prevenirse o corregirse con intervenciones tempranas y adecuadas.

Entre estas principales causas, el glaucoma, se estima que 80 millones padecen de esta enfermedad, se destaca como una de las enfermedades más desafiantes de diagnosticar en sus etapas iniciales. Esta afección se gana el sombrío apodo de ”ladrón silencioso de la visión”debido a que los pacientes no experimentan síntomas notables hasta que el daño ocular es bastante significativo, es por ello que aproximadamente el 50 % no saben que tienen glaucoma. Esta particularidad hace que la detección temprana del glaucoma sea un verdadero desafío para los profesionales médicos, ya que la enfermedad puede avanzar sin ser percibida por el paciente.

El glaucoma es una enfermedad ocular crónica que afecta el nervio óptico del ojo y puede llevar a la pérdida irreversible de la visión. Generalmente, se asocia con un aumento de la presión intraocular, que daña gradualmente las fibras del nervio óptico como se puede ver en la Figura 1. El tratamiento del glaucoma generalmente implica medicamentos para reducir la presión intraocular, cirugía láser o cirugía convencional, y su objetivo principal es controlar la progresión de la enfermedad y preservar la visión. Por lo tanto, el glaucoma requiere una atención médica constante y un seguimiento adecuado para evitar complicaciones visuales graves.

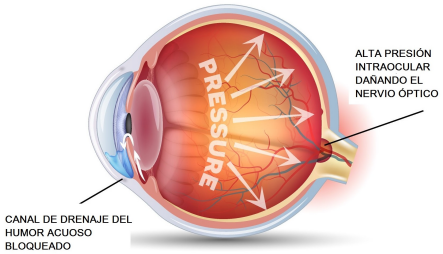

Figura 1. Glaucoma [13]

Sobre los problemas que trae el glaucoma nos indica [12] que la progresión del glaucoma conlleva una carga de síntomas tanto no visuales como visuales que son una preocupación considerable para los pacientes en casos más avanzados. Sin embargo, en el momento del diagnóstico, la mayoría de los pacientes están relativamente libres de discapacidades inducidas por el glaucoma. La ausencia de síntomas específicos para el glaucoma temprano parece contribuir a la alta cantidad de sujetos no diagnosticados encontrados en estudios de población. Los síntomas de la Escala del Glaucoma, como ardor, picazón, escozor, lagrimeo, sequedad, comezón, dolor y cansancio, pueden ocurrir en muchas enfermedades, como cualquiera de los síndromes de ojo seco, e incluso en ojos sanos durante la exposición a condiciones ambientales adversas. Sin embargo, en [12] nos indica que estos tipos de síntomas pueden ocurrir como consecuencia del tratamiento del glaucoma.

Evelyn C O’Neill et al. en [16] indicaron en el 2014 que “Si se muestra una imagen del nervio óptico a diferentes expertos, es probable que no estén de acuerdo en si es glaucomatoso o no, especialmente en las primeras etapas de la enfermedad”, continúa. “Los calificadores tienden a sobrestimar o subestimar el daño glaucomatoso y tienen baja reproducibilidad de calificación y poca

concordancia. Por lo tanto, entrenar modelos de IA para predecir clasificaciones subjetivas es problemático".

Sin embargo, en la actualidad esta enfermedad silenciosa ya logró ser detectada con la aplicación de la IA. En la búsqueda de soluciones a este problema de salud visual, se han llevado a cabo diversas investigaciones que emplean la IA. Estos estudios han aprovechado una variedad de técnicas de aprendizaje automático y clasificación de imágenes, utilizando extensas bases de datos de la enfermedad del glaucoma que incluyen recopilaciones de imágenes oftalmológicas y retinografías de pacientes con dicha enfermedad.

El resultado de estos esfuerzos es un sistema predictivo y de detección del glaucoma que ha arrojado resultados prometedores. La IA, siendo más específicos las CNN o redes convolucionales, se ha convertido en una herramienta invaluable para identificar indicios tempranos de la presencia del glaucoma, permitiendo a los médicos intervenir a tiempo y brindar el tratamiento necesario para preservar la visión del paciente.

1.1. Objetivos

Este artículo tiene como objetivo principal identificar las dos mejores técnicas de inteligencia artificial para la detección del glaucoma, a través del análisis de diversos trabajos que hacen uso de imágenes oftalmológicas y retinográficas. Además, se busca adquirir un profundo conocimiento sobre las bases de datos más utilizadas que contienen imágenes de pacientes afectados por esta enfermedad, con el propósito de entrenar clasificadores especializados en la detección del glaucoma. En última instancia, se pretende replicar proyectos similares para mejorar el dominio en el área, proponer una nueva idea implementando técnicas adicionales o mejoras a las existentes, y llevar a cabo una comparación exhaustiva de los resultados obtenidos.

2. Estado del Arte

En esta sección se llevará a cabo un análisis exhaustivo de diversos estudios que aportan al campo de las técnicas para la detección temprana del glaucoma. Se incluirá una tabla de síntesis para organizar y evaluar estos trabajos con el objetivo de identificar las contribuciones más destacadas.

2.1. Machine Learning

La introducción de la IA y el alto rendimiento computacional ha tenido un impacto significativo en el progreso de diversas disciplinas, incluyendo el diagnóstico automático de signos de glaucoma. Cuando se aplican adecuadamente, los modelos basados en algoritmos de aprendizaje automático pueden ofrecer diagnósticos más precisos en comparación con los métodos previos. En el artículo [21], se investiga una nueva estrategia basada en el algoritmo SVM para la clasificación automática de ojos con signos favorables a la patología del glaucoma. Durante la construcción de este modelo, se llevaron a cabo pruebas exhaustivas de varios parámetros de los algoritmos SVM, lo que resultó en un mejor rendimiento en las pruebas después de validar el modelo durante el entrenamiento. Este modelo, evaluado en términos de precisión, AUC y precisión, se ha demostrado tanto más confiable como eficiente en comparación con los enfoques anteriores para este problema, y también requiere menos tiempo de procesamiento.

En el 2022, Raul Huarote, et .a mostraron en [8] una investigación donde se cubria una necesidad de poder clasificar de acuerdo a los fondos de ojos en la enfermedad retinopatía diabética, glaucoma o que sean identificados como ojos sanos. Para lograr esto la estrategia particular está en cómo preparar las imágenes digitales en una secuencia, como convertir a tono de gris, realizar una ecualización, aplicar el algoritmo de resalte de borde canny y aplicar operaciones morfológicas para que se pueda ingresar a una red neuronal SOM (mapa autoorganizativo) y poder ser clasificada. Para lograr esto se tiene clasificada como 0 a retinopatía diabética, 1 a glaucoma y 3 a los ojos sanos. Para corroborar esta estrategia se ha toma d o una base de datos pública de Fundus-images, siendo 45 imágenes de ojos para el entrenamiento y para las pruebas se usaron 15 imágenes que no eran parte del entrenamiento y cada imagen en escala de grises está escalada a una dimensión de 256 x 256 pixeles, logrando demostrar con esta estrategia una afectividad de 93.7 % certeza en la identificación de clase de enfermedad ocular.

En el 2019 Mennato-Allah Talaa et al. en [18] mostraron que la diagnosis asistida por computadora (CAD) puede, por lo tanto, servir como un cambio significativo en la detección temprana del glaucoma al llevar al clínico al nivel de un experto. Además, CAD tiene la ventaja de ser no invasivo, simple y rentable. En este trabajo, se presenta un algoritmo genérico automatizado de detección de glaucoma en el que se calculan características estadísticas y texturales a partir de la región del disco óptico (ONH) en imágenes de la retina. Se realizaron varios análisis para comparar el rendimiento de la clasificación del glaucoma considerando diferentes técnicas de mejora de contraste (ecualización de histograma - ecualización de histograma adaptativa limitada de contraste) y modelos de color (RGB - HSV - CIELAB). Luego se utiliza la selección de características para encontrar el mejor conjunto de características para cada uno de los diferentes experimentos. El mejor rendimiento se logró cuando se calcularon características texturales a partir de los canales CIELAB con ecualización de histograma, lo que resultó en una precisión del 92.5 %, una sensibilidad del 95.0 % y una especificidad del 90.0 % considerando datos públicos, en este caso se usó la base de datos REFUGE.

Sin embargo, lo que se busca es poder detectar tempranamente el glaucoma antes de que esta enfermedad afecte gravemente al paciente. Es por ello que se pensó que podría darse otro enfoque a las imágenes con el glaucoma, es decir, enfocarse más en la retina y lo que pasaba con esta. Es así que en el artículo [6] se aprovecha que el glaucoma se indica tanto por cambios estructurales como por la presencia de atrofia en la retina. En imágenes de la retina, estos cambios se presentan en forma de variaciones sutiles en las intensidades locales. Estas variaciones suelen describirse mediante estadísticas basadas en la forma local, que son propensas a errores. Se propuso en [6] un enfoque automatizado basado en características globales para detectar el glaucoma en imágenes. Se ha diseñado una representación de la imagen para resaltar los indicadores sutiles de la enfermedad, de manera que las características globales de la imagen puedan discriminar de manera efectiva entre casos normales y casos de glaucoma.

El método propuesto en [6] se ha demostrado en un gran conjunto de imágenes anotadas por tres expertos médicos. Los resultados muestran que el método es eficaz en la detección de indicadores sutiles de glaucoma. El rendimiento de clasificación en un conjunto de datos de 1186 imágenes de retina en color, que contiene una mezcla de casos normales, casos sospechosos y casos confirmados de glaucoma, es del 97 por ciento de sensibilidad con un 87 por ciento de especificidad. Esto mejora aún más cuando se eliminan los casos sospechosos de los casos anormales. Por lo tanto, el método propuesto en [6] ofrece una solución efectiva para el cribado del glaucoma a partir de imágenes de la retina.

2.2. Redes Neuronales CNN

Se realizaron distintos proyectos relacionados con el uso de imágenes para detección la impresión del iris, es decir reconocer

si es real el iris. Lo importante a recalcar de este trabajo es el proceso para hacer este reconocimiento y la aplicación de la IA. Para esto Mehedi Hasan Raju, et al. en [17] proponen que la autenticación biométrica basada en el iris es una modalidad biométrica ampliamente utilizada debido a su precisión, entre otros beneficios. Mejorar la resistencia de la biometría del iris a los ataques de suplantación es un tema de investigación importante. El seguimiento ocular y los dispositivos de reconocimiento del iris tienen hardware similar que consta de una fuente de luz infrarroja y un sensor de imagen. Esta similitud potencialmente permite que los algoritmos de seguimiento ocular se ejecuten en sistemas de biometría impulsados por el iris. El trabajo actual avanza en el estado del arte de la detección de ataques de impresión del iris, en los que un impostor presenta una impresión del iris auténtico de un usuario a un sistema de biometría. La detección de estos ataques se logra mediante el análisis de la señal de movimiento ocular capturada con un modelo de aprendizaje profundo. Los resultados indican un mejor rendimiento del enfoque seleccionado en comparación con el estado del arte previo. En [17] se explica que se implementó una versión personalizada de la arquitectura ResNet 18 y la llamaron C-ResNet 18. En C-ResNet 18, se realizó varios cambios en comparación con ResNet 18 básico. Las razones principales detrás de estos cambios son adaptarla a la declaración del problema, conjunto de datos y lograr un mejor rendimiento en términos de métricas de evaluación. Cambiaron convoluciones 2D a 1D, se modificó la forma en que se usaban las conexiones de salto y se cambió el número de canales de entrada y salida en cada bloque de convolución. En resumen, C-ResNet 18 consta de 17 bloques de convolución, seguidos de un promedio global de pooling, una capa de aplanamiento y una capa lineal (totalmente conectada). Después de cada convolución, se aplicó normalización por lotes (BN) y la función de activación ReLU.

Abordando el problema de la aplicación de distintas técnicas apartir de imágenes del ojo para la detección del glaucoma, se recopilaron distintos artículos donde se analizó los trabajos y estudios realizados.

En 2017, los investigadores Daniel Shu .et al en [20] entrenaron un sistema de aprendizaje profundo para evaluar el glaucoma utilizando 125.189 imágenes de la retina. El rendimiento del algoritmo se validó en 71.896 imágenes. El resultado muestra una prevalencia del 0.1 % con un área bajo la curva para posible glaucoma de 0.942 (IC 95 %, 0.929 % a 0.954 %), una sensibilidad del 96.4 % (IC 95 %, 81.7 % a 99.9 %) y una especificidad del 87.2 % (IC 95 %, 86.8 % a 87.5 %).

En el 2020, fue publicado [4] donde se usó la base de datos RIM-ONE DL (RIM-ONE para Aprendizaje Profundo). En este artículo se describe este conjunto de imágenes, que consta de 313 retinografías de sujetos normales y 172 retinografías de pacientes con glaucoma. Todas estas imágenes han sido evaluadas por dos expertos e incluyen una segmentación manual del disco y la copa. También se describe un conjunto de pruebas de evaluación con diferentes modelos de redes neuronales convolucionales ampliamente conocidos. Donde se concluyó que el modelo de red VGG19 no solo proporcionó el AUC más alto, sino que su sensibilidad también igualó a 1, el valor más alto posible. El otro modelo de red con características similares, VGG16, también produjo buenos resultados. Aunque no es posible realizar una comparación directa con los resultados del desafío REFUGE, es interesante notar que el equipo ganador (Son et al., 2018) logró un AUC de 0.9885 con una sensibilidad del 0.9752 para una muestra de prueba que consistía en 360 imágenes de sujetos sanos y 40 imágenes de pacientes con glaucoma. Todo esto visto en [4]. El marco de evaluación propuesto en este trabajo contiene cuatro elementos principales: definición de los conjuntos de entrenamiento y prueba, modelos de redes neuronales utilizados, estrategia de entrenamiento y prueba empleada, y las métricas consideradas en la evaluación.

En lo que respecta a los conjuntos de entrenamiento y prueba utilizados se consideran dos variantes. En la primera variante, el conjunto de imágenes se dividió al azar en imágenes de entrenamiento y prueba utilizando una proporción de 70:30, respectivamente. En la segunda variante, las imágenes tomadas en el HUC se utilizaron para el entrenamiento (195 normales y 116 con glaucoma), y las imágenes tomadas en los otros dos hospitales se utilizaron para la prueba (118 normales y 56 con glaucoma). El único procesamiento realizado en las imágenes consistió en reescalarlas en intensidad en el rango de 0 a 1 y cambiar su tamaño a 224x224x3. Esto fue una de las cosas más resaltantes al ver el trabajo de [4].

Se pensó en realizar unas modificaciones en las imágenes, tales como la segmentación. Un proceso así se realizó en [22] donde Lianyi Wu, et al. descubrieron que la innovación clave fue la combinación SegNet y ADDA juntos. Se aplicó GAN para entrenar un codificador SegNet objetivo y combinandoló con un decodificador SegNet preentrenado. Este enfoque redujo el MSE (Error Cuadrático Medio) entre el CDR calculado con la segmentación y el CDR calculado con las anotaciones en el conjunto de datos objetivo. La contribución que trajo [22] también incluye experimentos sobre funciones de pérdida y experimentos sobre el tamaño del discriminador. Al mismo tiempo, se señaló las debilidades de ADDA en el artículo. Como trabajo a futuro de este artículo fue que se probarían más enfoques de aprendizaje por transferencia y se utilizaría alguna técnica para ampliar el conjunto de datos de entrenamiento.

Otro trabajo que obtuvo buenos resultados y probó con arquitecturas anteriormente mencionadas fue realizado por Andres Diaz-Pinto, et al. donde en su articulo [7] se analizaron cinco arquitecturas de redes neuronales convolucionales (CNN) entrenadas en ImageNet (VGG16, VGG19, InceptionV3, ResNet50 y Xception) y se utilizaron como clasificadores de glaucoma. Utilizando solo bases de datos disponibles públicamente, la arquitectura Xception mostró el mejor rendimiento para la clasificación de glaucoma, lo cual se evaluó como el equilibrio entre el AUC y el número de parámetros de la CNN. Basándose en 1707 imágenes y una técnica de aumento de datos, se obtuvo un AUC promedio de 0.9605 con un intervalo de confianza del 95 % de 95.92-97.07 %, una especificidad promedio de 0.8580 y una sensibilidad promedio de 0.9346 después de afinar la arquitectura Xception, mejorando significativamente otros trabajos de última generación.

Además, un análisis adicional muestra que el modelo afinado tiene un rendimiento competitivo cuando se prueba en imágenes que provienen de una base de datos completamente diferente. Este experimento difiere del enfoque común en el que un subconjunto de una base de datos se utiliza para entrenamiento y el otro subconjunto se utiliza para pruebas. Utilizando la base de datos ACRIMA como conjunto de prueba únicamente, se obtuvo un AUC de 0.7678 con un intervalo de confianza del 95 % de 68.41-81.81 %. El mismo experimento se realizó en las otras cuatro bases de datos públicas: HRF, Drishti-GS1, RIM-ONE, sjchoi86-HRF, obteniendo un AUC de 0.8354, 0.8041, 0.8575 y 0.7739, respectivamente.

La base de datos más relevante en [7] fue ACRIMA, la cual está compuesta por 396 imágenes de glaucoma y 309 imágenes normales y podría ser fácilmente utilizada como banco de pruebas para comparaciones y/o análisis adicionales. Los autores alientan a la comunidad científica a probar sus modelos utilizando la nueva base de datos disponible públicamente y comparar sus resultados con el método propuesto en este artículo.

2.3. Enfoque de la retinografía

En el artículo [11] publicado en el 2018, se vió que mediante la detección de la relación copa-disco (la proporción del área de la copa óptica al área del disco óptico, CDR) de los pacientes, se puede realizar un cribado del glaucoma. Utilizando imágenes de retina de alta resolución en la base de datos HRF, se propone un conjunto completo de métodos de cribado. En primer lugar, el realiza la extracción de canales y la mejora de la imagen, luego utiliza un algoritmo de segmentación por umbral para separar el disco óptico y utiliza el método de crecimiento de regiones para separar la copa óptica. De esta manera, se puede calcular automáticamente el CDR. El [11] utiliza 30 imágenes para realizar pruebas, 15 de ellas son imágenes de ojos normales y las otras 15 son de ojos con glaucoma. Los resultados mostraron que los CDR de los pacientes con glaucoma son mayores de 0,6 en todas las 15 imágenes de glaucoma, y los CDR de los ojos normales están entre 0,2 y 0,6. Esto concuerda con el juicio de los expertos.

Después de analizar los distintos trabajos presentados en está sección, se realizó una tabla para poder resaltar los datos más relevantes de estos trabajos como: la técnica o arquitectura que utilizaron, en lo que más se enfocaron para utilizar en sus pruebas, la base de datos y la confiablididad obtenida. Todo esto se puede observar en el Cuadro 1.

3. Marco Teórico

Este trabajo explora las redes neuronales en el contexto de la inteligencia artificial, destacando su capacidad para realizar aprendizaje automático y reconocimiento de patrones. Estas redes, inspiradas en el funcionamiento del cerebro humano, emplean capas interconectadas de neuronas artificiales que procesan información de manera distribuida. Las redes neuronales profundas, o *Deep Learning*, con múltiples capas ocultas, han revolucionado la inteligencia artificial al aprender representaciones jerárquicas de datos. Su versatilidad las hace cruciales en aplicaciones que van desde el reconocimiento de imágenes hasta el diagnóstico médico, transformando la interacción con la tecnología y abordando desafíos en diversas disciplinas.

En este trabajo se habla sobre la aplicación de la redes convolucionales, por ello es importante entender que las redes convolucionales, también conocidas como CNN, son un tipo especializado de arquitectura de redes neuronales profundas diseñadas específicamente para el procesamiento y análisis de datos visuales, como imágenes y videos. Estas redes son altamente efectivas en la extracción automática de características relevantes de las imágenes a través de capas de convolución y pooling, lo que les permite identificar patrones visuales complejos, como bordes, texturas y formas, de manera jerárquica. Las redes convolucionales se utilizan en una amplia gama de aplicaciones de visión por computadora, incluyendo la clasificación de imágenes, la detección de objetos, el reconocimiento facial y la segmentación de imágenes, y han demostrado un rendimiento sobresaliente en tareas de procesamiento de imágenes y videos en la actualidad.

Para la extracción de estas características relevantes de los datos de entrada, como una imagen, mediante la aplicación de filtros o núcleos de convolución a través de la imagen se utilizan las operaciones convolucionales. Estos filtros son matrices pequeñas que se deslizan o convolucionan sobre la imagen de entrada. Cada filtro detecta patrones o características específicas en la imagen, como bordes, texturas o formas.

El proceso de convolución implica los siguientes pasos:

1. Seleccionar un filtro o núcleo de convolución, que es una matriz de pesos.
2. Colocar el filtro en una ubicación específica de la imagen de entrada.
3. Realizar una multiplicación elemento por elemento entre los valores del filtro y los valores correspondientes en la región de la imagen cubierta por el filtro.
4. Sumar los productos resultantes para obtener un solo valor, que representa la respuesta del filtro en esa ubicación.
5. Mover el filtro a lo largo de la imagen de entrada realizando el mismo cálculo en diferentes ubicaciones.
6. Crear una nueva imagen llamada ”mapa de características.º ”mapa de activación”que contiene las respuestas del filtro en cada ubicación.

Estos mapas de características resultantes pueden pasar por otras capas de la red neuronal, como capas de agrupación y capas completamente conectadas, para realizar tareas específicas, como clasificación de objetos en imágenes.

Las operaciones convolucionales son efectivas para capturar patrones locales en los datos de entrada, lo que las hace particularmente útiles en tareas de visión por computadora, como detección de objetos, reconocimiento de imágenes y segmentación semántica. También reducen la cantidad de parámetros en la red en comparación con las redes neuronales completamente conectadas, lo que ayuda a evitar el sobreajuste y permite el aprendizaje eficiente de características relevantes.

Se puede ver la arquitectura básica de estas CNN en la Figura 2.

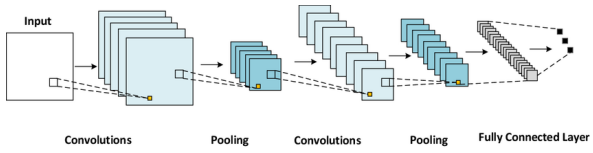

Figura 2. Arquitectura CNN [1]

La función SoftMax es una función de activación que se utiliza en la capa de salida de una red neuronal para realizar la clasificación multiclase. Esta función se utiliza para convertir las salidas de la capa anterior en probabilidades que suman uno. Las probabilidades se utilizan para medir la confianza del modelo en la pertenencia a cada clase.

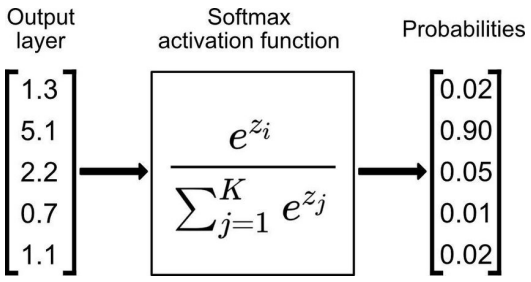

Figura 3. Función de activación Softmax [5]

Esta toma como entrada un vector de valores y produce una salida en forma de vector de la misma dimensión. La función SoftMax realiza dos operaciones: primero, exponencia los elementos del vector de entrada, y segundo, divide cada uno de los elementos exponenciados por la suma de los elementos exponenciados, como podemos ver en la Figura 3.

Esta función se utiliza en la clasificación multiclase porque asigna una probabilidad a cada clase. Esto significa que es útil

Ref	IA	Enfoque	Técnica	Base de datos	AUC
[21]	ML	-	SVM	Glaucoma Center of Semmelweis University Budapest	89.47 %,
[17]	CNN	Impresión de iris	C-ResNet 18	ETPAD v2	87.78 %
[20]	DLS	Retinografías	-	SIDRP	94.2 %
[8]	UML	-	SOM	Fundus-Images	93.7 %
[18]	ML	Retinografías	SVM, RGB - HSV - CIELAB	REFUGE	92.5 %
[4]	CNN	Retinografías	VGG19	RIM-ONE - REFUGE	92.72 %
[11]	-	CDR	Segmentation	HRF	98.74 %
[22]	CNN	CDR	SegNet - ADDA	-	-
[6]	ML	DBL	GMP	-	97 %
[7]	CNN	Retinografías	Xception	RIM-ONE, ACRIMA, HRF, Drishti-GS1, sjchoi86-HRF	96.05 %

Cuadro 1. REF: REFERENCIA, AUC: ÁREA BAJO LA CURVA ROC, ML: *Machine Learning*, DLS: APRENDIZAJE PROFUNDO SUPERVISADO, CDR: RADIO DE LA COPA A DISCO, GMP: PATRÓN DE MOMENTO GENERALIZADO, DBL: DETECCIÓN DE LESIONES BRILLANTES, SVM: *Support Vector Machine*

cuando se necesita saber la probabilidad de que una instancia pertenezca a cada clase. De igual manera utiliza en muchas aplicaciones, como la clasificación de imágenes, la clasificación de texto y la clasificación de voz.

Tiene diversas ventajas esta función en el proceso de clasificación. En primer lugar, convierte las salidas de la capa anterior en probabilidades que suman uno, lo que facilita la interpretación de los resultados. En segundo lugar, se puede utilizar para clasificación multiclase. Sin embargo, la función SoftMax tiene algunas limitaciones. La función SoftMax no es adecuada para problemas de clasificación binaria, y tiende a exagerar las diferencias entre las probabilidades.

En resumen, la función SoftMax es una función de activación importante en la clasificación multiclase. Esta función se utiliza para convertir las salidas de la capa anterior en probabilidades que suman uno. La función SoftMax se utiliza en aplicaciones como la clasificación de imágenes y la clasificación de texto. La función SoftMax tiene ventajas, como la facilidad de interpretación de los resultados, y limitaciones, como la falta de adecuación para problemas de clasificación binaria. En general, la función SoftMax es una herramienta fundamental en la clasificación multiclase y debe ser considerada en cualquier problema de clasificación.

El Max-Pooling es un proceso de discretización basado en muestras. Su objetivo es submuestrear una representación de entrada (imagen, matriz de salida de capa escondida, etc.) reduciendo su tamaño. Además, su interés es que reduce el coste de cálculo reduciendo el número de parámetros que tiene que aprender y proporciona una invariancia por pequeñas translaciones (si una pequeña translación no modifica el máximo de la región barrida, el máximo de cada región seguirá siendo el mismo y por tanto la nueva matriz creada será idéntica).

Para que la acción del Max-Pooling sea más concreta, se puede ver en el ejemplo mostrado en la Figura 4 donde imaginaremos que tenemos una matriz de 4×4 que representará la entrada inicial y un filtro de una ventana de 2×2 que se aplicará en nuestra entrada. Para cada región barrida por el filtro, el Max-pooling tomará el máximo y de ese modo creará a la vez una nueva matriz de salida en la que cada elemento corresponderá a los máximos de cada región detectada.

Es importante hablar de la función de activación ReLU (*Rectified Linear Unit*), la cual es una función ampliamente utilizada en las CNN y en otros tipos de redes neuronales artificiales. Su función principal es introducir no linealidad en la red, lo que permite que la red pueda aprender y modelar relaciones más complejas en los datos.

La función ReLU se define de la siguiente manera:

$ReLU(x) = max(0, x)$

Donde x es la entrada a la función. La función ReLU toma un valor de cero para cualquier entrada negativa y pasa directamente

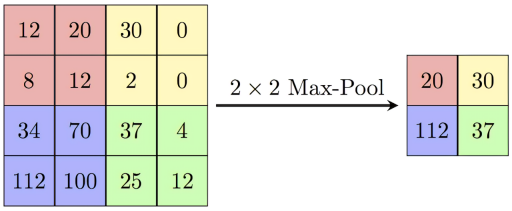

Figura 4. Procesamiento Max-Pooling de 4x4 a 2x2 [3]

cualquier entrada positiva sin modificarla. En otras palabras, si la entrada es positiva, la función ReLU la deja pasar tal como es; si es negativa, la función ReLU la transforma en cero. Esto crea una activación esparsa en la red, ya que solo las neuronas cuya entrada es positiva se activan.

En la arquitectura Xception se aplica el mecanismo de residuos, también conocido como conexiones residuales, es una técnica clave en las CNN que permite que la información fluya a través de capas convolucionales de manera más eficiente. En lugar de simplemente apilar capas una encima de la otra, las conexiones residuales introducen conexiones directas desde una capa a otra, permitiendo que la información se ”salte.ª través de capas. Esto se logra mediante la suma de la salida de una capa a la salida de la capa subsiguiente como se puede ver en la Figura 5. En esencia, se agrega una especie de ruta corta”que permite que los gradientes se propaguen más fácilmente durante el entrenamiento. Esto ayuda a abordar el problema de la degradación del rendimiento que puede ocurrir en redes muy profundas, donde agregar capas adicionales puede disminuir la precisión del modelo. Las conexiones residuales permiten entrenar con éxito redes extremadamente profundas y han sido un componente clave en el desarrollo de arquitecturas de vanguardia como ResNet.

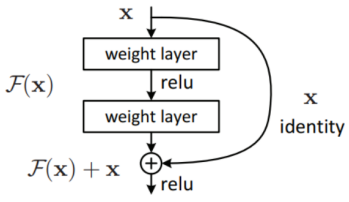

Figura 5. Red Residual [2]

Finalmente, la validación cruzada de 5 pliegues, a menudo abreviada como ”validación cruzada de 5-fold.º çross-validation de 5-fold”, es una técnica comúnmente utilizada en el aprendizaje

automático y la evaluación de modelos. Esta técnica se emplea para evaluar la capacidad de generalización de un modelo y reducir el sesgo en la evaluación del rendimiento del modelo. Esta técnica consiste en:

1. División de datos: Supongamos que tienes un conjunto de datos que deseas utilizar para entrenar y evaluar un modelo. En el caso que mencionaste, se está utilizando un conjunto de datos para clasificar imágenes de glaucoma y normal.
2. Partición en pliegues (*folds*): En la validación cruzada de 5 pliegues, el conjunto de datos se divide en 5 subconjuntos o pliegues aproximadamente del mismo tamaño.
3. Ciclo de entrenamiento y evaluación: Se realiza un ciclo de entrenamiento y evaluación 5 veces, correspondiente al número de pliegues. En cada ciclo:
   - Pliegue de prueba: Un pliegue se reserva como conjunto de prueba.
   - Pliegues de entrenamiento: Los otros 4 pliegues se utilizan como conjunto de entrenamiento.
   - Entrenamiento y evaluación: Se entrena el modelo en los pliegues de entrenamiento y se evalúa su rendimiento en el pliegue de prueba. Esto implica ajustar el modelo a los datos de entrenamiento y medir su precisión en los datos de prueba.
4. Promedio de resultados: Después de completar los 5 ciclos, se obtienen 5 medidas de rendimiento (una para cada pliegue de prueba). Estas medidas suelen ser métricas como precisión, exactitud, sensibilidad, especificidad, entre otras, dependiendo de la tarea. Para evaluar el rendimiento general del modelo, se calcula el promedio de estas medidas.

La validación cruzada de 5 pliegues permite una evaluación más robusta y confiable del modelo, ya que se evalúa en diferentes subconjuntos de datos. También ayuda a evitar el sobreajuste, ya que se prueba en múltiples particiones de los datos. Además, se utiliza para determinar la estabilidad del rendimiento del modelo en diferentes subconjuntos de datos, lo que es esencial en tareas médicas, como la clasificación de imágenes de glaucoma y normal, donde la variabilidad en los datos puede ser significativa.

4. Técnicas a implementar

A través de los trabajos relacionados se observó que las técnicas de clasificación respecto a las imágenes de ojos con la enfermedad con glaucoma con mejores resultados fueron Xception, [7] en donde alcanzó un 96.32 % de AUC, y VGG19, [4] en donde alcanzó un 92.72 % de AUC, estos datos pueden observarse en el Cuadro 1. Para obtener dichos resultados comparativos se aplicó el AUC que significa "Área Bajo la Curva ROC", la cual es una métrica utilizada para evaluar el rendimiento de modelos de clasificación, incluyendo las redes neuronales convolucionales (CNN). Mide la capacidad de un modelo para distinguir entre clases positivas y negativas a través de diferentes valores de umbral. El valor de AUC varía de 0 a 1, siendo un valor más alto indicativo de un mejor rendimiento del modelo.

En [19] se tiene una lista de las distintas arquitecturas que son aplicadas en distintos proyectos como clasificadores de imágenes, en esta tabla mostrada en la Figura 6 la precisión top-1 y top-5 se refiere al rendimiento del modelo en el conjunto de datos de ImageNet. La profundidad se refiere a la profundidad topológica de la red. Esto incluye capas de activación, capas de normalización

Model	Size (MB)	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth	Time (ms) per inference step (CPU)	Time (ms) per inference step (GPU)
Xception	88	79.0%	94.5%	22.9M	81	109.4	8.1
VGG16	528	71.3%	90.1%	138.4M	16	69.5	4.2
VGG19	549	71.3%	90.0%	143.7M	19	84.8	4.4
ResNet50	98	74.9%	92.1%	25.6M	107	58.2	4.6
ResNet50V2	98	76.0%	93.0%	25.6M	103	45.6	4.4
ResNet101	171	76.4%	92.8%	44.7M	209	89.6	5.2

Figura 6. Modelos disponibles, Aplicaciones con Keras [19]

por lotes, etc. El tiempo por paso de inferencia es el promedio de 30 lotes y 10 repeticiones.

En esta sección se hará una descripción de dichos clasificadores y una comparación de estos.

4.1. Xception

Se desarrolló para abordar algunas de las limitaciones de las arquitecturas convencionales, como VGG o ResNet, y se basa en un concepto llamado "factorización en bloques de convolución"para mejorar la eficiencia y el rendimiento. La arquitectura de este clasificador se puede observar en la Figura 7, en esta arquitectura cabe resaltar algunas partes importantes de esta:

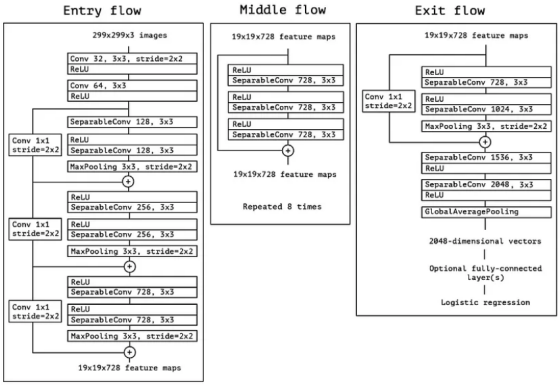

Figura 7. Arquitectura General de Xception (Flujo de entrada > Flujo medio > Flujo de salida) [14]

- Compuesta por 133 capas convolucionales.
- Bloques de convolución separables: En una convolución separable, se dividen en dos etapas: primero se aplican convoluciones 1x1 independientes para mezclar canales y luego se aplican convoluciones 3x3 para realizar convoluciones espaciales. Esto reduce significativamente el número de parámetros y la carga computacional.
- Reducción de dimensiones: Xception utiliza el max-pooling, para reducir las dimensiones espaciales de las características a medida que avanzan a través de la red.
- Función de activación ReLU: Al igual que en muchas otras arquitecturas de CNN, Xception utiliza la función de activación Rectified Linear Unit (ReLU) después de cada capa convolucional para introducir no linealidad en la red.
- Capas completamente conectadas: Las capas combinan las características aprendidas en las capas convolucionales y aplican una función softmax para calcular las probabilidades de pertenencia a cada clase.

- Aplicación de mecanismos residuales después de cada capa convolucional.

Es por ello que Xception es una arquitectura de CNN innovadora que se destaca por su eficiencia y rendimiento en tareas de clasificación de imágenes. Su enfoque en bloques de convolución separables le permite mantener un alto nivel de capacidad de representación mientras reduce la cantidad de recursos computacionales requeridos, lo que la hace valiosa en aplicaciones donde la eficiencia es esencial.

En [7], la arquitectura Xception y otras más se ajustaron finamente para la tarea de evaluación de glaucoma utilizando sus versiones entrenadas en ImageNet disponibles en el núcleo de Keras. Para utilizar estas redes en esta tarea, se cambió la última capa completamente conectada de cada CNN por una capa de agrupación global promedio (*GlobalAveragePooling2D*), seguida de una capa completamente conectada con dos nodos que representan dos clases (glaucoma y saludable) y un clasificador softmax. Por lo tanto, contando las nuevas capas superiores en cada CNN, el número total de capas de Keras en las arquitecturas de ResNet50 y Xception están compuestas por 176 y 133 capas de Keras, respectivamente. Es importante recalcar que las imágenes se recortaron automáticamente alrededor del disco óptico.

Con el fin de obtener el mejor rendimiento de cada modelo, se realizaron varios experimentos variando el número de capas ajustadas finamente y el número de épocas.

Primero, en cuanto al número de capas, se ajustó finamente la última capa ponderada de las arquitecturas de CNN, manteniendo las otras capas en modo ”no entrenable”. Luego, se aumentó el número de capas ajustadas finamente hasta actualizar todas las capas en la CNN.

El segundo experimento consistió en analizar el impacto del número de épocas que presentan el mejor rendimiento para cada arquitectura. Otros hiperparámetros como el tamaño del lote, la tasa de aprendizaje, etc., se mantuvieron fijos mientras se varió el número de capas ajustadas finamente y el número de épocas.

Además de estos experimentos, también se evaluó el rendimiento de las CNN utilizando la técnica de validación cruzada k-fold con k = 10, siguiendo el procedimiento descrito en [9]. Para evitar el sobreajuste y aumentar la robustez de los modelos, se aumentaron las imágenes disponibles mediante rotaciones aleatorias, zoom en un rango entre 0 y 0.2 y volteos horizontales y verticales. Las imágenes también se redimensionaron al tamaño de entrada predeterminado de cada arquitectura de CNN (299 × 299 para Inceptionv3 y Xception).

Se llevó a cabo una evaluación de rendimiento particular de las CNN, utilizando conjuntos de datos que no se utilizaron durante la etapa de entrenamiento. A diferencia de la mayoría de los trabajos en la literatura, este experimento verificó el rendimiento de las CNN en bases de datos completas que el sistema no ha visto durante la etapa de entrenamiento.

El último experimento es la comparación entre la mejor de las cinco CNN que se utilizaron y un algoritmo de última generación que también utiliza bases de datos públicas.

Las imágenes utilizadas se enfocan en el disco óptico, copa óptica y el área de la retina, esto se puede ver en la Figura 8

4.2. VGG19

VGG19 es una arquitectura de red neuronal convolucional que se utiliza comúnmente como clasificador de imágenes. Fue desarrollada por el *Visual Geometry Group* (VGG) en la Universidad de Oxford y es una extensión de la arquitectura VGG16. VGG19 es conocida por su profundidad, ya que consta de 19

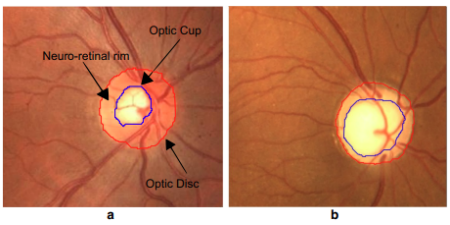

Figura 8. Imágenes digitales del fondo de ojo recortadas alrededor del disco óptico. a Estructuras principales de un disco óptico sano y b Disco óptico glaucomatoso [9].

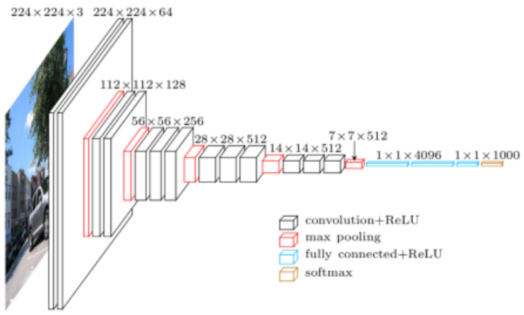

Figura 9. Arquitectura de la red VGG19 [10].

capas de convolución y pooling, 16 capas convolucionales y 3 capas totalmente conectadas.

A continuación, explicaré las principales características de la arquitectura VGG19 como clasificador de imágenes:

- Se proporcionó una imagen RGB de tamaño fijo de (224 * 224) píxeles como entrada a esta red, lo que significa que la matriz tenía una forma de (224, 224, 3).
- El único preprocesamiento realizado fue restar el valor medio RGB de cada píxel, calculado sobre todo el conjunto de entrenamiento.
- Utiliza relleno espacial (spatial padding) para preservar la resolución espacial de la imagen.
- Usa una operación de agrupación máxima (max pooling) sobre ventanas de 2 * 2 píxeles con un paso (stride) de 2.
- Esto fue seguido por la unidad lineal rectificada (ReLU) para introducir no linealidad y mejorar la capacidad de clasificación del modelo, así como para mejorar el tiempo de cálculo, ya que los modelos anteriores utilizaban funciones como tangente hiperbólica (tanh) o sigmoides, lo que demostró ser mucho mejor que esas.
- Se implementaron tres capas completamente conectadas, de las cuales las dos primeras tenían un tamaño de 4096, y después de eso, una capa con 1000 canales para la clasificación de 1000 categorías en el conjunto de datos *Imagenet Large Scale Visual Recognition Challenge* (ILSVRC), y la capa final es una función SoftMax.

Es por ello que VGG19 es una arquitectura robusta y eficiente para tareas de clasificación de imágenes debido a su profundidad y capacidad para aprender características jerárquicas de las imágenes. Sin embargo, su principal inconveniente es su tamaño y complejidad, lo que hace que su entrenamiento sea más lento y requiera más recursos computacionales. Se puede ver la arquitectura del VGG19 en la Figura 9.

En el [4] se utilizaron retinografías, donde se observa el fondo del ojo, un ejemplo se puede observar en la Figura 10, donde

se han destacado las partes más relevantes para diagnosticar el glaucoma: el disco óptico, la copa y el borde neurorretiniano. Para esta arquitectura se tomó como entrenamiento la base de datos RIM-ONE, mencionada anteriormente en nuestro estado del arte.

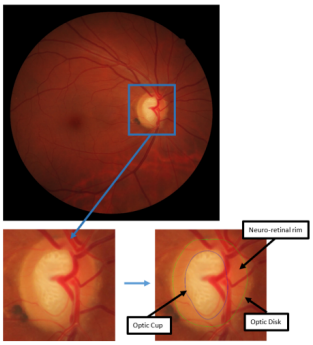

Figura 10. Muestra de retinografía con las regiones más relevantes para el diagnóstico del glaucoma [4].

En cuanto a los modelos de redes neuronales utilizados, se probaron la mayoría de las arquitecturas contenidas en el Framework de Aprendizaje Profundo Keras: Xception, VGG16, VGG19, ResNet50, InceptionV3, InceptionResNetV2, MobileNet, DenseNet121, NASNetMobile y MobileNetV2. Donde el clasificador VGG19 fué el más óptimo. En cada caso, el tamaño de la capa de entrada se estableció en 224x224x3, y se añadió una capa *GlobalAveragePooling2D* a la base convolucional, seguida de una capa de salida totalmente conectada con dos salidas, utilizando SoftMax para distinguir entre las clases Normal y Glaucoma.

La estrategia de entrenamiento fue la misma para ambas variantes de los conjuntos de datos. Comenzamos con las redes preentrenadas utilizando los valores de pesos de ImageNet proporcionados por Keras y se ajustaron finamente todas las capas. [7] encontraron que esto producía los mejores resultados. Para evitar el sobreajuste, se utilizó la ampliación de datos que consiste en rotaciones aleatorias (-30°, 30°), volteo vertical y horizontal, y zoom (0.8, 1.2).

Los pasos realizados para ajustar finamente las redes fueron los siguientes:

1. Congelar la base de la red convolucional.
2. Entrenar la parte que se añadió.
3. Descongelar todas las capas en la base de la red.
4. Entrenar conjuntamente todas las capas en la red.

Para aumentar la confiabilidad de los experimentos, se aplicó una validación cruzada de 5 pliegues, con una proporción del 80 % para el conjunto de entrenamiento y el 20 % para el conjunto de validación en cada pliegue. Para cada pliegue, se siguieron los pasos enumerados para determinar el número de épocas más adecuado en los pasos 2 y 4. Para el entrenamiento en el paso 2, se utilizó un tamaño de lote de 32, junto con un optimizador RMSprop con una tasa de aprendizaje de 2e-5 y la entropía cruzada categórica como función de pérdida. Para el entrenamiento en el paso 4, la tasa de aprendizaje se estableció en 1e-5. Una vez completada la fase de validación, el modelo final para cada red se entrenó utilizando todos los datos de entrenamiento (sin divisiones) y siguiendo los mismos cuatro pasos que antes para el número de épocas que maximizó la precisión promedio de validación en los pasos 2 y 4. Este modelo final se utilizó para evaluar cada red en el conjunto de prueba.

En lo que respecta a las métricas utilizadas, se aplicó el enfoque descrito en [15], en el que se utiliza el AUC como medida de evaluación de referencia. Esta medida se complementó con el valor de sensibilidad (Se = T p / (T p + Fn)) a una especificidad de 0.85 (Sp = T n / (T n + F p)), donde T p, F p, T n y Fn representan el número de casos verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos, respectivamente. Esto permite evaluar el rendimiento de las diferentes redes cuando se desea minimizar la tasa de falsos positivos. La tercera medida incluida es la exactitud, que es común en este tipo de problemas, aunque se reconoce que puede mostrar sesgo en conjuntos de datos con clases desequilibradas.

Finalmente se hicieron distintas pruebas en [4], donde una de ellas reflejó que después de hacer una prueba con datos de Madrid y Zaragoza, las cuales componen el *dataset* RIM-ONE, que el VGG19 tuvo la mayor curva de aprendizaje y mejores resultados, esto se puede observar en la Figura 11.

Network	AUC	Se	Acc.
VGG19	0.9272	0.8750	0.8563
VGG16	0.9177	0.8214	0.8506
InceptionV3	0.9015	0.7500	0.8046
Xception	0.8982	0.7500	0.7989
DenseNet	0.8919	0.7143	0.7816
MobileNet	0.8912	0.7500	0.8276
ResNet50	0.8855	0.7321	0.8333
InceptionResNetV2	0.8396	0.625	0.7644
NASNetMobile	0.7969	0.6071	0.7989
MobileNetV2	0.7765	0.4464	0.5287

Figura 11. Evaluación de las distintas redes utilizando el conjunto de prueba de RIM-ONE [7].

Al analizar estos dato se ve una disminución en las otras arquitecturas. Esta disminución podría explicarse por el hecho de que se utilizó un conjunto de imágenes de prueba cuya apariencia visual era bastante diferente de la de las imágenes utilizadas durante el entrenamiento. Es importante tener en cuenta que las imágenes se capturaron en diferentes hospitales en circunstancias diferentes, lo que parece haber afectado a las redes. La falta de robustez de este tipo de sistema ante distorsiones que pueden afectar a las imágenes, como el ruido, el contraste o la iluminación, ha sido analizada por varios autores. Por ello es importante siempre recalcar estos incidentes.

4.3. Comparación de técnicas

En esta sección se hará una comparación en aspectos específicos entre ambas técnicas.

- Profundidad de la arquitectura: Xception es una arquitectura de red profunda basada en el concepto de separable convolutions (convoluciones separables). Tiene una profundidad significativa, pero utiliza bloques de convoluciones separables en lugar de convoluciones estándar, lo que reduce el número de parámetros y, en teoría, mejora la eficiencia de la red. Por otro lado, VGG19 es parte de la familia de redes VGG, y "19.[en] su nombre hace referencia a que tiene 19 capas, lo que lo convierte en una red profunda pero más simple en comparación con algunas arquitecturas más modernas. Utiliza convoluciones estándar.
- Parámetros y tamaño del modelo: Debido al uso de separable convolutions, Xception tiene menos parámetros en comparación con otras arquitecturas profundas, lo que lo hace más eficiente en términos de almacenamiento y

cómputo. En cambio, VGG19 tiene un mayor número de parámetros debido a su profundidad y uso de convoluciones estándar. Esto lo hace más grande en términos de tamaño de modelo y requerimientos de memoria.

- Rendimiento: Xception ha demostrado ser eficaz en una variedad de tareas de visión por computadora y reconocimiento de imágenes, con un buen equilibrio entre precisión y eficiencia. De igual manera VGG19 es conocido por su capacidad de aprendizaje profundo y su alto rendimiento en tareas de clasificación de imágenes, pero a costa de tener más parámetros y ser más costoso en términos computacionales.
- Usos comunes: Debido a su eficiencia y buen rendimiento, Xception es a menudo preferido en aplicaciones donde se requiere un buen rendimiento pero con restricciones de recursos, como en dispositivos móviles o sistemas embebidos. Mientras que VGG19 se usa comúnmente en la investigación y desarrollo de algoritmos de visión por computadora y como base para transferencia de aprendizaje en tareas específicas de clasificación de imágenes.

5. Experimentos y resultados

En esta sección se describirá la base de datos que se utilizó en ambas técnicas. Se explicarán los experimentos realizados en ambas técnicas para alcanzar el *accuracy* de las técnicas analizadas en la sección anterior, se comentarán las dificultades al momento de la ejecución de los experimentos. Finalmente, se analizarán los resultados obtenidos en los experimentos. Las pruebas y evidencias de lo obtenido se puede observar en el https://github.com/RodATS/Proyecto_Carrera.git.

5.1. Base de datos: RIM-ONE

RIM-ONE DL es un conjunto de datos integral de imágenes retinianas diseñado específicamente para evaluar el glaucoma mediante el uso de técnicas de aprendizaje profundo. Este conjunto de datos unificado presenta 313 retinografías de sujetos normales y 172 retinografías de pacientes diagnosticados con glaucoma, de las cuales se dividirán en: 217 imágenes de entrenamiento, 174 imágenes para el test y 94 imágenes para la validación. Las imágenes fueron adquiridas en tres destacados hospitales españoles: el Hospital Universitario de Canarias en Tenerife, el Hospital Universitario Miguel Servet en Zaragoza, y el Hospital Clínico Universitario San Carlos en Madrid, [4].

Para garantizar la consistencia y la comparabilidad de los resultados entre diversas publicaciones, se recomienda utilizar las particiones originales del conjunto de datos para fines de entrenamiento y prueba. Es esencial destacar que se insta a los usuarios a emplear únicamente los datos proporcionados por RIM-ONE DL, sin agregar imágenes adicionales de otras bases de datos al entrenar modelos o ajustar algoritmos. La calidad y la diversidad de las retinografías, junto con la procedencia de múltiples instituciones médicas de prestigio, hacen de RIM-ONE DL un recurso valioso para el avance en la detección y evaluación del glaucoma mediante enfoques basados en el aprendizaje profundo.

5.2. Experimento técnica Xception y resultados

Para la técnica de Xception, el artículo [7] tiene un drive donde se tienen los archivos necesarios para realizar las pruebas. Tiene un archivo donde indica como organizar las carpetas para comenzar con los experimentos. En [7] se utilizaban varias bases de datos, es por ello que se realizó la modificación para sólo trabajar con RIM-ONE.

5.2.1. Entorno de ejecución. Para la réplica de la técnica de Xception se utilizó:

- Procesador: Intel(R) Core(TM) i5-10210U CPU @ 1.60GHz 2.11 GHz
- RAM: 12.0 GB (11.7 GB utilizable)
- GPU: Nvidia Force MX250
- Memoria GPU: 8GB

5.2.2. Primer experimento y resultado. El script tiene como objetivo evaluar la estabilidad y los intervalos de confianza de las puntuaciones AUC de un modelo de aprendizaje profundo en la base de datos RIM-ONE utilizando un enfoque de remuestreo bootstrap. Los resultados se guardan en un archivo CSV y se visualiza a través de un histograma, el cual se puede observar en la Figura 12. Para estas pruebas se realizaron 100 iteraciones y se analizaron las imágenes para los tests.

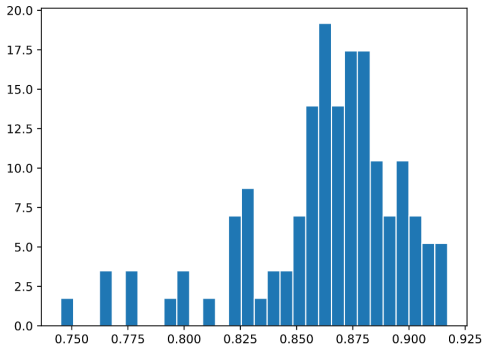

Figura 12. Histograma para evaluación y calcular el AUC

Después de realizar la prueba de confiabilidad utilizando la base de datos RIM-ONE, se obtuvo como resultado un 95 % de AUC y 93.89 % de *accuracy*, mientras en [7] se obtuvo 96.05 % de AUC y 89.77 % de *accuracy*.

5.2.3. Segundo experimento y resultado. Se realizaron unas pruebas para analizar diferentes imágenes de retinografías y pasar por el modelo entrenado para detectar si se presenta glaucoma o no. Para estas pruebas se utilizó la base de datos pública ACRIMA, disponible en el archivo que brinda [7]. Algunas de las imágenes utilizadas se pueden observar en las Figuras 13 y 14.

Figura 13. Im310_g_ACRIMA

Estos resultados se pueden ver en el Cuadro 2.

5.3. Experimentos técnica VGG19 y resultados

En el artículo [4] no daban un código para realizar las pruebas de la CNN. Sin embargo, nos brinda la base de datos, las segmentaciones para las imágenes y los pesos para que usó para el entrenamiento de la arquitectura.

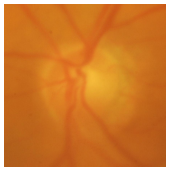

Figura 14. Im311_g_ACRIMA

Nombre Imagen	Glaucoma	No glaucoma
Im310_g_ACRIMA.jpg	0.61027	0.38973
Im311_g_ACRIMA.jpg	0.68036	0.31964
Im312_g_ACRIMA.jpg	0.79831	0.20169
Im313_g_ACRIMA.jpg	0.99628	0.00372

Cuadro 2. TEST DE GLAUCOMA, DATA SET ACRIMA.

5.3.1. Entorno de ejecución. Se utilizó en está réplica el Google Collab, las características son:

- Modelo: Intel(R) Xeon(R) CPU @ 2.20GHz
- Familia de CPU: 6
- Modelo: 79
- Velocidad del procesador: 2.20 GHz
- Núcleos de CPU: 1 (por núcleo virtual)
- Hilos (threads): 2
- Tamaño de la caché: 56320 KB
- Arquitectura: 46 bits físicos, 48 bits virtuales

5.3.2. Primeros experimentos y resultados. Para estos primeros experimentos se realizó con la CNN VGG19 que nos brinda Keras. Y en estos experimentos se fueron aplicando paulatinamente los cambios a las imágenes, como por ejemplo la aplicación de las segementaciones, es decir, máscaras.

- **En el Experimento 1:** se hace un entrenamiento de 224 x 224 x 3, con el globalAverage2D, como indica el paper y se realizá el entrenamiento. Nos da como *accuracy*: 69.78 %.
- **En el Experimento 2:** se hace un entrenamiento de 224 x 224 x 3, con el globalAverage2D y se aplicaron a las imágenes rotaciones aleatorias (-30°, 30°), volteo vertical y horizontal, y zoom (0.8, 1.2), como indica el paper y se realizá el entrenamiento. Nos da como *accuracy*: 69.78 %.
- **En el Experimento 3:** Se aplicó lo mismo que el Entrenamiento 2 y se añadirá las máscaras a las imágenes. Nos da como *accuracy*: 67.82 %.

5.3.3. Siguientes experimentos y resultados. En estos experimentos se modificó la CNN VGG19 según la estructura que nos menciona en [4]. Recordando algunas de las especificaciones que se mencionaron anteriormente: tamaño de entrada de 224x224x3; una capa de Global Average Pooling 2D; que la capa de salida sea una capa densa con dos unidades y una activación softmax, lo cual es consistente con la especificación de una salida completamente conectada con dos salidas para distinguir entre las clases ”Normalz ”Glaucoma”; utilizar la estrategia de ajuste fino, iniciando con pesos preentrenados de ImageNet y afinando todas las capas en cuatro etapas: congelación de la base convolucional, entrenamiento de la parte añadida, descongelación de todas las capas y entrenamiento conjunto de todas las capas; a los datos aplicarles las rotaciones aleatorias, volteos verticales y horizontales, y zoom para evitar el sobreajuste.

- **En el Experimento 1:** después de modificar la arquitectura VGG19 y usar los pesos que nos brinda el artículo, se realizaron las pruebas correspondienttes y dió como *accuracy*: 67.82 %.
- **En el Experimento 2:** Se optó por diseñar el modelo utilizando la arquitectura de la CNN VGG19, aunque con ajustes específicos no detallados en el artículo. Se introdujo la capa de Dropout antes de la capa densa para regularizar las conexiones antes de la capa de salida. Además, se aseguró que la configuración del optimizador y la tasa de aprendizaje coincidiera con las especificaciones, utilizando el optimizador RMSprop con una tasa de aprendizaje de 2e-5. En el paso 4 del ajuste fino, la tasa de aprendizaje se ajustó a 1e-5. Se llevó a cabo una revisión exhaustiva de las etapas de ajuste fino para garantizar su implementación correcta en el código. Tras realizar el experimento, se logró un *accuracy* del 71.76 % y un AUC de 58.88 %.

En el artículo [4] se alcanzó el 85 % de *Accuracy* y un AUC de 92.72 % en las pruebas, así que por la escases de información no se llegó al *accuracy* y AUC esperado. Sin embargo, se logró dominar mejor las CNN y el data set RIM-ONE, es importante recalcar que se logró aplicar las segmentaciones o máscaras a las imágenes, lo cual era una de las partes más relevantes de esta técnica ya que se centraba únicamente en el Disco Óptico y la Copa Óptica, las cuales mencionadas anteriormente son factores importantes para la detección del glaucoma debido a las consecuencia de las variaciones de sus dimensionalidades durante la enfermedad. Se puede ver en las Figuras 15, 16 y 17 un ejemplo de la aplicación de las segmentaciones o máscaras que nos brindaban en el respositorio de [4] a las imágenes del *data set* RIM-ONE.

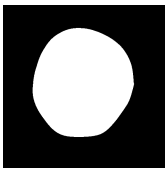

Figura 15. Máscara para el Disco Óptico, RIM-ONE.

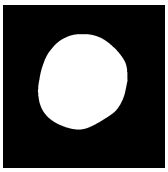

Figura 16. Máscara para la Copa Óptica, RIM-ONE.

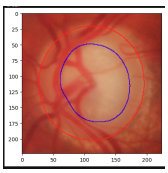

Figura 17. Aplicación de la máscara de la Copa y Disco Óptico, RIM-ONE.

Para la implementación de las máscaras, el artículo [4] brinda los archivos los cuales se observan en las Figuras 15 y 16, de igual manera contiene en su repositorio un código para aplicar las máscaras, las cuales después de varios intentos y correcciones no lograba funcionar correctamente. Es por ello que se optó en realizar

una función para la aplicación de estas, los resultados se pueden observar en la Figura 17.

6. Conclusión

Como primer objetivo de este artículo científico era conocer las mejores técnicas para la detección del glaucoma. Como se logró observar en este trabajo, el glaucoma a pesar de ser una de las enfermedades más difíciles de detectar a principios de su aparición se entendió que aplicando un clasificador de imágenes puede hacerse está detección teniendo buenos resultados. Después de una investigación y comparación de trabajos se encontró que los clasificadores que obtuvieron los mejores resultados fueron Xception donde en uno de los trabajos alcanzó 96.05 % de AUC y 89.77 % de *accuracy*, y la arquitectura VGG19 la cual alcanzó un 92.72 % de AUC y 85 % de *accuracy* mostrado en otro trabajo. Para la evalución y comparación de técnicas se enfatizó en el AUC, "Area Bajo la Curva ROC", la cual es una métrica utilizada para evaluar el rendimiento de modelos de clasificación, y el *accuracy* o confiablidad, la cual indica la probabilidad de la buena detección del glaucoma, teniendo un margen bajo de error. De igual manera se observaron las ventajas y desventajas de cada una de estas arquitecturas entendiendo que si se busca eficiencia y buen rendimiento en dispositivos con recursos limitados, Xception puede ser la mejor opción; si existe acceso a una gran cantidad de recursos computacionales y estás enfocado en tareas de clasificación de imágenes de alta calidad, VGG19 puede ser una elección sólida. Después de realizar los experimentos con la primera técnica, Xception, para la detección de glaucoma se alcanzó un AUC de 95 % y *accuracy* de 93.89 %, siendo una réplica exitosa. Por otro lado, en la técnica VGG19 se alcanzó un AUC de 58.88 % y *accuracy* de 71.76 %, debido a ciertas dificultadas mencionadas anteriormente no se lograron los valores esperados. Finalmente, se aprendió como estos clasificadores haciendole pequeñas variaciones a sus arquitecturas y a los datos de entrada, en este caso la base de datos RIM-ONE, puede ayudar a la detección del glaucoma en sus inicios y es de está manera que brindan un aporte a la medicina.

Referencias

[1] A convolutional neural network.

[2] Figure-1. ResNet Architecture [10]. (A) difference between a plain net...

[3] Max-pooling / Pooling - Computer Science Wiki.

[4] Francisco José Fumero Batista, Tinguaro Diaz-Aleman, Jose Sigut, Silvia Alayon, Rafael Arnay, and Denisse Angel-Pereira. Rim-one dl: A unified retinal image database for assessing glaucoma using deep learning. *Image Analysis Stereology*, 39(3):161–167, 2020.

[5] BotPenguin. Softmax Function: Advantages and Applications — BotPenguin, 11 2023.

[6] K. Sai Deepak, Madhulika Jain, Gopal Datt Joshi, and Jayanthi Sivaswamy. Motion pattern-based image features for glaucoma detection from retinal images. In *Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing*, ICVGIP '12, New York, NY, USA, 2012. Association for Computing Machinery.

[7] Andrés Diaz-Pinto, Sandra Morales, Valery Naranjo, Thomas Köhler, José M. Mossi, and Amparo Navea. CNNs for Automatic Glaucoma assessment using FundUs Images: An extensive validation. *Biomedical Engineering Online*, 18(1), 3 2019.

[8] Huarote Zegarra Raúl Eduardo. .Estrategia para la detección de tipos de enfermedades oculares usando Red Neuronal SOM , 7 2022.

[9] Trevor Hastie, Robert Tibshirani, and Jerome H. Friedman. *The elements of statistical learning*. 1 2009.

[10] Redacción KeepCoding. Arquitectura VGG16 y VGG19 en Deep learning, 11 2023.

[11] Zhenyu Liu, Yakun Gao, and Sicheng Zhu. Research of screening method based on glaucoma image. In *Proceedings of the 3rd International Conference on Multimedia and Image Processing*, ICMIP '18, page 114–118, New York, NY, USA, 2018. Association for Computing Machinery.

[12] Charles W. McMonnies. Glaucoma history and risk factors. *Journal of Optometry*, 10(2):71–78, 4 2017.

[13] Miranza. Glaucoma ocular ¿Qué es y cómo se produce? — Miranza, 8 2021.

[14] Sabbir Mohammed. Https://maelfabien.github.io/deeplearning/xception/ — Data Science blog, 6 2019.

[15] José Ignacio Orlando, Huazhu Fu, João Barbossa Breda, Karel Van Keer, Deepti R. Bathula, Andrés Diaz-Pinto, Ruogu Fang, Pheng-Ann Heng, Je Young Kim, Joon-Ho Lee, Xiaoxiao Li, Peng Liu, Shiping Lu, Balamurali Murugesan, Valery Naranjo, Sai Samarth R Phaye, Sharath M Shankaranarayana, Apoorva Sikka, Jaemin Son, Anton Van Den Hengel, Shujun Wang, Junyan Wu, Zifeng Wu, Guanghui Xu, Yongli Xu, Pengshuai Yin, Fēi Li, Xiulan Zhang, Yanwu Xu, and Hrvoje Bogunović. REFUGE Challenge: A Unified Framework for evaluating Automated Methods for Glaucoma Assessment from FundUs Photographs. *Medical Image Analysis*, 59:101570, 1 2020.

[16] Evelyn C O'Neill, Lulu U. Gurria, Surinder Singh Pandav, Yu Kong, Jessica Brennan, Jing Xie, Michael Coote, and Jonathan G Crowston. Glaucomatous Optic Neuropathy Evaluation Project. *JAMA Ophthalmology*, 132(5):560, 5 2014.

[17] Mehedi Hasan Raju, Dillon J Lohr, and Oleg Komogortsev. Iris print attack detection using eye movement signals. In *2022 Symposium on Eye Tracking Research and Applications*, ETRA '22, New York, NY, USA, 2022. Association for Computing Machinery.

[18] Mennato-Allah Talaat, Nataly Raed, Aya Medhat, Romisaa Ashraf, Mohammad Essam, Rana ElKashlan, and Lamiaa Abdel-Hamid. Glaucoma Detection from Retinal Images using Generic Features. *ACM Association for Computing Machinery*, 9 2019.

[19] Keras Team. Keras documentation: Keras applications, no date.

[20] Daniel Shu Wei Ting, Carol Yim-Lui Cheung, Gilbert Lim, Gavin Siew Wei Tan, Nguyen D. Quang, Alfred Gan, Haslina Hamzah, Renata Garcia-Franco, Ian Yew San Yeo, Shu Yen Lee, Edmund Yick Mun Wong, Charumathi Sabanayagam, Mani Baskaran, Farah Ibrahim, Ngiap Chuan Tan, Eric A. Finkelstein, Ecosse L. Lamoureux, Ian Y. Wong, Neil M. Bressler, Sobha Sivaprasad, Rohit Varma, Jost B. Jonas, Ming Guang He, Ching-Yu Cheng, Gemmy Chui Ming Cheung, Tin Aung, Wynne Hsu, Mong Li Lee, and Tien Yin Wong. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *JAMA*, 318(22):2211–2223, 12 2017.

[21] Stéphane Cédric Koumétio Tékouabou, El Arbi Abdellaoui Alaoui, Imane Chabbar, Walid Cherif, and Hassan Satori. Machine Learning Aprroach for Early Detection of Glaucoma from Visual Fields. *ACM Association for Computing Machinery*, 3 2020.

[22] Lianyi Wu, Yiming Liu, Yelin Shi, Bin Sheng, Ping Li, Lei Bi, and Jinman Kim. Detect glaucoma with image segmentation and transfer learning. In *Proceedings of the 32nd International Conference on Computer Animation and Social Agents*, CASA '19, page 37–40, New York, NY, USA, 2019. Association for Computing Machinery.