Aplicación de MobileNet para el diagnóstico temprano del glaucoma: Un enfoque binocular

O. Kovalyk Borodyak [1], J. Morales Sánchez [1], R. Verdú Monedero [1], I. Sellés Navarro [2], J.L. Sancho Gómez [1]

[1] Universidad Politécnica de Cartagena, Cartagena, España,
olekesandr.kovalyk@edu.upct.es, {juan.morales, rafael.verdu, josel.sancho}@upct.es
[2] 2 Hospital Universitario Reina Sofía, Murcia, España, inmasell@um.es

Resumen

Este estudio presenta los resultados preliminares de la detección temprana del glaucoma utilizando la red neuronal MobileNet. MobileNet es una red eficiente en términos del uso de memoria y recursos computacionales. En este trabajo se comparan dos enfoques: monocular, que considera un ojo, y binocular, que integra información de ambos ojos. La investigación utiliza la base de datos PAPILA, con retinografías y datos clínicos de 244 pacientes. Los resultados muestran que el enfoque binocular supera al monocular, con un aumento del 9 % en el área bajo la curva (AUC) y un 15 % en sensibilidad a una especificidad del 90 %. Estos resultados indican una posible ventaja del uso de ambos ojos de un paciente para la mejora del poder de diagnóstico de las redes neuronales.

1. Introducción

El glaucoma es la principal causa de ceguera irreversible en todo el mundo [1]. Esta patología suele aparecer en edades avanzadas, generalmente causada por alta presión intraocular, afecta la cabeza del nervio óptico (CNO) y provoca una pérdida progresiva del campo visual. Se sabe que los indicadores del glaucoma aparecen años antes que las lesiones que causan pérdida del campo visual. En este contexto, la detección precoz es de vital importancia para minimizar los daños producidos por la enfermedad y los costes sanitarios [2]. Las principales pruebas que se realizan para detección del glaucoma son: la tomografía de coherencia óptica, la tonometría, la campimetría y la retinografía. La retinografía (Figura 1) es una imagen del fondo ocular, normalmente centrada en el disco óptico, que sirve para el propósito de estudiar la morfología de la CNO. Cabe destacar que la retinografía es una modalidad de imagen médica minimamente invasiva.

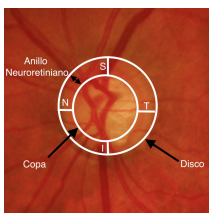

Figura 1: *Datos morfológicos de una retinografía. Se han marcado las longitudes Inferior (I), Superior (S), Nasal (N) y temporal (T).*

Con el avance de la tecnología en medicina, y particularmente con la irrupción de las técnicas de procesamiento de imágenes y la inteligencia artificial, las redes neuronales convolucionales han ganado un protagonismo considerable en el diagnóstico de diversas patologías, incluido el glaucoma. Estas herramientas, se caracterizan por su capacidad para analizar grandes volúmenes de imágenes médicas, ofreciendo una respuesta en cuestión de segundos.

Por otro lado, la hipótesis sobre la asimetría entre las características anatómicas de la CNO de ambos ojos ha sido recientemente reconsiderada como un indicador para el diagnóstico temprano del glaucoma [3, 4, 5, 6]. Discrepancias significativas entre ambos ojos de un mismo paciente pueden ser indicativas de alteraciones patológicas. En este sentido, las técnicas avanzadas de aprendizaje automático podrían ofrecer la posibilidad de aprovechar estas diferencias para la detección temprana del glaucoma.

Dentro de este marco de trabajo, proponemos un método basado en la arquitectura MobileNet [7] para el diagnóstico automático del glaucoma. Se ha escogido esta red en particular dado que puede ser implementada en dispositivos compactos, como retinógrafos portátiles, simultáneamente, permitiría analizar la morfología del nervio óptico, eludiendo así las demoras que pudieran surgir debido a la necesidad de acceder a un retinógrafo tradicional. Esto facilitaría a los profesionales de la atención primaria realizar evaluaciones inmediatas para descartar la presencia de glaucoma. El modelo se evaluará considerando dos enfoques: el enfoque monocular y el enfoque binocular. En el enfoque binocular información de ambos ojos de forma simultánea, accediendo así a características a las que de otro modo no son accesibles. Los resultados del enfoque binocular se compararán con el modelo monocular, el cual predice el glaucoma a partir de una única imagen. Con todo ello se pretende diseñar una herramienta efectiva y eficiente para la detección temprana del glaucoma, mejorando así el pronostico de los pacientes afectados por esta enfermedad.

2. Materiales

2.1. Conjunto de datos

En este trabajo, emplearemos la base de datos PAPILA [8], la cual es de acceso abierto, proporciona retinografías de ambos ojos por paciente, segmentaciones realizadas por dos expertos y datos clínicos. Elegimos PAPILA por diversos motivos: al ser de acceso abierto facilita la replicabilidad de resultados, ofrece información de ambos ojos de cada paciente, está estructurada como una muestra representativa de la población e in-

ISBN: 978-84-17853-76-1

cluye el grado de progresión del glaucoma. PAPILA comprende datos de ambos ojos de 244 pacientes, es decir un total de 488 retinografías. Éstas se distribuyen en 333 etiquetadas como sanas, 87 diagnosticadas con glaucoma y 68 consideradas sospechosas de glaucoma.

2.2 Preparación del conjunto de datos

En primer lugar, se excluyeron de las pruebas los pacientes considerados por los expertos como sospechosos de padecer glaucoma. Los paciente sospechosos son aquellos cuyas pruebas no son concluyentes como para diagnosticarlos. En pruebas anteriores se ha observado que la etiqueta sospechosa no es una etiqueta intermedia, si no que algunos de los sospechosos realmente tienen glaucoma y otros no. Esto provocaba un impacto negativo en el entrenamiento del algoritmo. Por ello las etiquetas de pacientes sospechosos del glaucoma se han descartado para este estudio.

Debido al tamaño relativamente pequeño del conjunto de datos, se hace uso de técnicas de validación cruzada, concretamente el método *k-fold*. Este método consiste en dividir el conjunto de datos en k subconjuntos llamados *folds*. En nuestra investigación se ha considerado $k = 5$. Cada uno de estos *folds* sirve como conjunto de validación, mientras que el resto se emplea para el entrenamiento. En este trabajo en concreto se ha considerado dos particiones k-fold, una para el ajuste de hiperparámetros y otra para la presentación de resultados finales.

Para la segmentación de datos, adoptamos criterios específicos: la división se basa en pacientes y no en ojos individuales y se tiene en cuenta el nivel de progresión del glaucoma. Identificamos tres niveles de progresión: glaucoma temprano, glaucoma medio y glaucoma severo. En el modelo monocular, se descarta aleatoriamente la imagen de un ojo por paciente. Así, tanto en el modelo binocular como en el monocular contamos con 42 instancias para test y 168 para entrenamiento.

Antes del entrenamiento, las imágenes se normalizaron en un rango de 0 a 1 dividiendo las imágenes entere 255. Se ha observado en experimentos anteriores que los tiempos de entrenamiento se reducen de manera significativa al reducir el tamaño de las imágenes, sin que esto afecte de manera significativa a los resultados. Por lo que para este estudio las imágenes se han redimensionado a $224 \times 224 \times 3$. Durante el entrenamiento, se aplicaron técnicas de aumento de datos, con el fin de conseguir un modelo más robusto y obtener mejores resultados. Los aumentos de datos se han realizado de manera aleatoria (50 % por para aplicar cada transformación) durante el entrenamiento y son los siguientes: volteo horizontal y vertical, rotación de 0 a 180 grados, transformación afín, transformación de la perspectiva, desenfoque Gaussiano y la alteración de color.

3. Métodos

3.1. MobileNet monocular

MobileNet[7] es una arquitectura de red neuronal convolucional diseñada para ser eficiente y ligera. Para conseguir esto los autores de MobileNet hacen uso de: convoluciones separables por profundidad, multiplicadores de ancho y resolución y eficiencia en parámetros. En las versiones *v2* y *v3* de MobileNet se introducen funciones de activación más eficientes tales como *ReLU6* y *h-swish*

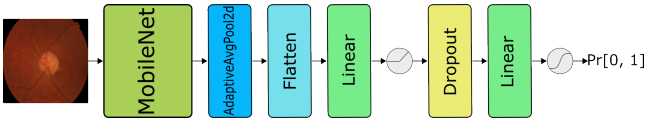

Figura 2: *Esquema general del método con el enfoque monocular.*

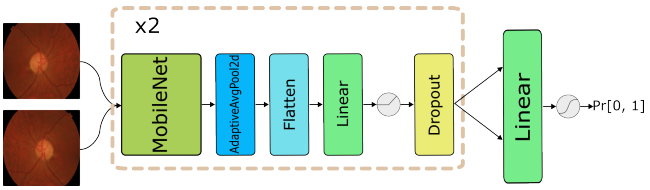

Figura 3: *Esquema general del método con el enfoque binocular.*

Arquitecturas como VGG, ResNet u otras arquitecturas que pueden ofrecer un rendimiento ligeramente superior en términos de precisión, pero también son mucho más pesadas en cuanto al número de parámetros y operaciones de cómputo. MobileNet, por otro lado, esta diseñada con un equilibrio entre precisión y eficiencia, abriendo así la posibilidad de su utilización en dispositivos con recursos limitados, tales como smartphones, dispositivos móviles o dispositivos IoT.

En la Figura 2 podemos ver un esquema general del algoritmo en modo monocular. En la entrada tenemos una retinografía que tras el preprocesado mencionado anteriormente se le suministra a MobileNet. La red extrae las características obteniendo así un espacio latente. Este espacio latente se filtra pasándolo por una capa *Adaptive Average Pooling* y posteriormente se aplana con una capa *flatten*, obteniendo así el vector latente. Tras ello, el vector resultante se administra a la capa lineal, a la cual le sigue de una función de activación ReLU6. Con el objetivo de evitar el sobreajuste se hace uso de una capa *dropout* (solo durante el entrenamiento). Finalmente, la capa de salida con activación *Softmax* devuelve la probabilidad de que una imagen tenga glaucoma o no.

3.2. MobileNet binocular

La Figura 3 muestra un esquema del algoritmo en su modo monocular. En el enfoque binocular la estructura del algoritmo es prácticamente la misma salvo por algunas modificaciones. La entrada pasa a ser dos imágenes. De estas imágenes se extrae un vector latente único que contiene información de ambos ojos. Este vector latente es el que se le administra a la capa lineal para establecer el diagnóstico.

3.3. MobileNet implementación

La implementación de la red está realizada en lenguaje Python, concretamente con la librería de aprendizaje

ISBN: 978-84-17853-76-1

máquina PyTorch, de la mano de TorchVision [9]. Las versión de las librería de PyTorch y TorchVision son 2.0.1 y 0.15.2 respectivamente. La versión de MobileNet que se ha utilizado es la *v3*.

3.4. Entrenamiento

Para ambos modos, binocular y monocular, la red se configuró de la siguiente manera: una tasa de aprendizaje establecida en $8,08 \times 10^{-5}$, un *weight decay* de $5,35 \times 10^{-6}$, y un tamaño de lote de 2. Se asignó un peso a cada clase basado en la frecuencia inversa de la misma. Las primeras 20 capas de la MobileNet se mantuvieron congeladas y se utilizaron pesos preentrenados de ImageNet [10]. Además, se aplicó suavizado de etiquetas (*label smoothing*) con un valor de 0,003 durante 30 épocas. La elección de los distintos valores de los hiperparámetros corresponden a las pruebas realizadas sobre un conjunto *k-fold* distinto al conjunto sobre el cual se han presentado los resultados. Los valores que se han elegido son aquellos que devolvían un menor error entre los 5 *folds*.

3.5. Hardware

Los experimentos se han realizado en un servidor de cómputo con sistema Ubuntu Linux 20.04.06 LTS, instalado sobre el siguiente hardware: procesador AMD EPIC 7643 de 48 núcleos, RAM de 256 GB y tarjeta gráfica NVIDIA 3090 de 24GB de VRAM.

4. Resultados y discusión

En esta sección, analizaremos los resultados obtenidos con el método propuesto, enfocándonos en su capacidad diagnóstica. Utilizaremos tres métricas ampliamente conocidas en el análisis de rendimiento de algoritmos de aprendizaje automático, especialmente en el ámbito del diagnóstico médico: sensibilidad, especificidad y el área bajo la curva, del acrónimo AUC en inglés. La sensibilidad computa el tanto por ciento (o tanto por uno) de pacientes que se clasifican como enfermos siendo estos realmente enfermos. La especificidad lo que computa es el tanto por ciento de los paciente diagnosticados como sanos de los que realmente lo son. En este trabajo, como en otros del mismo estilo [11], se fijará la especificidad al 80, 85 y 90 por ciento y se medirá la sensibilidad obtenida. Además, se destacará la eficiencia del método en términos de tiempo de cálculo.

4.1. Capacidad de diagnóstico

Los resultados obtenidos bajo el enfoque monocular se presentan en la Tabla 1 y la Figura 4. La Tabla 1 muestra la sensibilidad alcanzada para especificidades fijadas al 80 %, 85 % y 90 %. Esto se hace para obtener un valor en tanto por cien de enfermos que se detectan preservando las cantidades fijadas de tanto por cien de sanos detectados como tal. En lugar de promediar los resultados de cada *fold*, el total, se calcula uniendo en un único conjunto, los conjuntos (*folds*) de test de todos los 5 *folds*.

A pesar de las dificultades inherentes al conjunto de datos, que representa situaciones clínicas realistas, la sensibilidad para una especificidad del 80 % en cada pliegue se mantiene por encima del 50 %. En general, se logra una sensibilidad superior al 50 % para las tres especificidades mencionadas y un área bajo la curva (AUC) del 81 %, como se ilustra en la Figura 4. Al igual que para el total en la Tabla 1 y 2, el cálculo de la curva AUC se realiza tomando los datos de los 5 *folds* juntos.

Sensib. a la Especif. de	80 %	85 %	90 %
Fold 1	60	60	30
Fold 2	50	50	20
Fold 3	57,14	57,14	57,14
Fold 4	50	50	37,5
Fold 5	66,66	33,33	33,33
Total	**65,9**	**63,63**	**54,54**

Tabla 1: *Resultados considerando un solo ojo de cada paciente.*

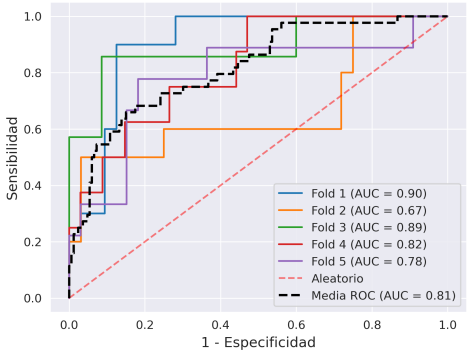

Figura 4: *Curva Característica de Operación del Receptor para los 5 folds y el total con el enfoque monocular.*

En la Tabla 2 podemos observar los resultados obtenidos al tener en cuenta ambos ojos de un mismo paciente. Se puede observar una notoria mejora en la sensibilidad. Esta supera el 66 % para especificidades fijadas al 80 % y 85 % en todos los *folds*. observando el total de la Tabla 2, la sensibilidad excede el 70 % para cada valor de especificidad. Además, el área bajo la curva (AUC), ilustrada en la Figura 5, alcanza el 90 % al considerar ambos ojos a la misma vez.

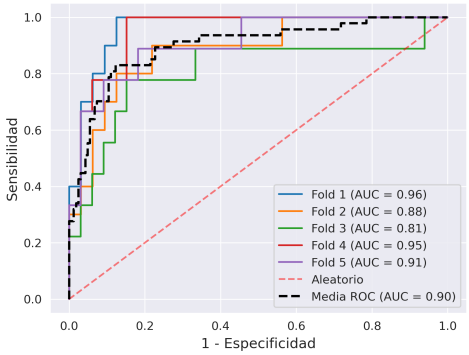

Figura 5: *Curva Característica de Operación del Receptor para los 5 folds y el total con el enfoque binocular.*

ISBN: 978-84-17853-76-1

Sensib. a la Especif. de	80 %	85 %	90 %
Fold 1	90	90	80
Fold 2	80	70	60
Fold 3	66,6	66,6	44,4
Fold 4	77,7	77,7	66,6
Fold 5	77,7	77,7	66,6
Total	**82,97**	**80,85**	**70,21**

Tabla 2: *Resultados considerando ambos ojos de cada paciente.*

Como muestran las Tablas 1 y 2, además de las figuras 4 y 5, la incorporación de información de ambos ojos ha conducido a una mejora significativa en los resultados. En términos cuantitativos, esta mejora se traduce en incrementos del 17,07 %, 17,22 % y 15,67 % en la sensibilidad para especificidades del 80 %, 85 % y 90 %, respectivamente. Además, se observa un aumento del 9 % en la métrica de AUC promedio.

4.2. Tiempo de cómputo

La Tabla 3 muestra los tiempos de cálculo por pliegue. Es evidente que al tener en cuenta la información de ambos ojos el tiempo de cálculo aumenta ligeramente ($\times 1,3$), tanto en la fase de entrenamiento como en la de inferencia. Hay que remarcar que en la fase de inferencia el algoritmo es capaz de procesar 46 imágenes (o 92 si se consideran ambos ojos) en menos de un segundo. Esta eficiencia lo posiciona como un método altamente viable para aplicaciones en tiempo real en dispositivos con recursos limitados.

Tiempo de	Entrenamiento	Inferencia
Un ojo	$80,32 \pm 1,051$	$0,65 \pm 0,02$
Ambos ojos	$96{,}64 \pm 0{,}44$	$0{,}86 \pm 0{,}03$

Tabla 3: *Tiempo promedio de ejecución entre folds del método propuesto en segundos (media $\pm$ desviación).*

5. Conclusiones

En este trabajo se ha explorado la influencia de considerar la información de ambos ojos en una red diseñada para el diagnóstico automático de glaucoma con recursos limitados. Los resultados indican que al tener en cuenta ambos ojos de un mismo paciente se logra una mejora significativa: un incremento del 9 % en el AUC y un aumento de más del 15 % en la sensibilidad para una especificidad del 90 %.

Este hallazgo sugiere que la consideración conjunta de ambos ojos podría ser valiosa en la detección del glaucoma en el marco del aprendizaje máquina. Una posible razón podría ser que, al procesar simultáneamente la información latente de ambos ojos, la red puede acceder a características morfológicas que de otro modo pasarían inadvertidas. Sin embargo, dado que estos son resultados preliminares, aún no es posible establecer conclusiones definitivas al respecto, por lo que se precisa de un análisis exhaustivo y pruebas complementarias de validación.

Referencias

[1] Yahya Shaikh, Fei Yu, and Anne L. Coleman. Burden of undetected and untreated glaucoma in the united states. *American Journal of Ophthalmology*, 158(6):1121–1129.e1, 2014.

[2] Alfred Sommer, Joanne Katz, Harry A Quigley, Neil R Miller, Alan L Robin, Ronald C Richter, and Kathe A Witt. Clinically detectable nerve fiber atrophy precedes the onset of glaucomatous field loss. *Archives of ophthalmology*, 109(1):77–83, 1991.

[3] Rafael Berenguer-Vidal, Rafael Verdú-Monedero, Juan Morales-Sánchez, Inmaculada Sellés-Navarro, and Oleksandr Kovalyk. Analysis of the asymmetry in rnfl thickness using spectralis oct measurements in healthy and glaucoma patients, 2022.

[4] Tahereh Mahmudi, Raheleh Kafieh, Hossein Rabbani, Alireza Mehri, and Mohammad-Reza Akhlaghi. Evaluation of asymmetry in right and left eyes of normal individuals using extracted features from optical coherence tomography and fundus images. *Journal of Medical Signals and Sensors*, 11(1):12–23, 2021.

[5] Donald L. Budenz. Symmetry between the right and left eyes of the normal retinal nerve fiber layer measured with optical coherence tomography (an aos thesis). *Transactions of the American Ophthalmological Society*, 106:252–275, 2008.

[6] L. S. Ong, P. Mitchell, P. R. Healey, and R. G. Cumming. Asymmetry in optic disc parameters: the blue mountains eye study. *Investigative Ophthalmology & Visual Science*, 40(5):849–857, Apr 1999.

[7] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications, 2017.

[8] Oleksandr Kovalyk, Juan Morales-Sánchez, Rafael Verdú-Monedero, Inmaculada Sellés-Navarro, Ana Palazón-Cabanes, and José-Luis Sancho-Gómez. Papila: Dataset with fundus images and clinical data of both eyes of the same patient for glaucoma assessment. *Scientific Data*, 9(1):1–12, 2022.

[9] TorchVision maintainers and contributors. Torchvision: Pytorch's computer vision library, 2016.

[10] Jia Deng, R. Socher, Li Fei-Fei, Wei Dong, Kai Li, and Li-Jia Li. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, volume 00, pages 248–255, 06 2009.

[11] Rui Fan, Kamran Alipour, Christopher Bowd, Mark Christopher, Nicole Brye, James A Proudfoot, Michael H Goldbaum, Akram Belghith, Christopher A Girkin, Massimo A Fazio, Jeffrey M Liebmann, Robert N Weinreb, Michael Pazzani, David Kriegman, and Linda M Zangwill. Detecting glaucoma from fundus photographs using deep learning without convolutions: Transformer for improved generalization. *Ophthalmology science*, 3(1):100233, March 2023.