Un enfoque explicativo para el diagnóstico de la degeneración macular asociada a la edad mediante técnicas de deep learning

M. Herrero-Tudela[1], R. Romero-Oraá[1,2], R. Hornero Sánchez[1,2], Gonzalo C. Gutiérrez-Tobal[1,2], M. I. López Gálvez[1,2], M. García[1,2]

[1] Grupo de Ingeniería Biomédica, Universidad de Valladolid, Valladolid, España,

{maria.herrero.tudela, roberto.romero, roberto.hornero, gonzalocesar.gutierrez, maria.garcia.gadanon}@uva.es

[2] Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), España

Resumen

La degeneración macular asociada a la edad (DMAE) es un trastorno que afecta a la mácula, una zona de la retina clave para la agudeza visual. La DMAE es una de las causas más frecuentes de ceguera en personas mayores de 60 años en los países desarrollados. Aunque se han propuesto tratamientos que frenan su desarrollo, su eficacia disminuye significativamente en las fases avanzadas. Por ello, son importantes los programas de cribado para la detección precoz. Sin embargo, implementar tales programas para enfermedades como la DMAE suele ser inviable debido a la extensa población de riesgo y a la necesidad de que los profesionales revisen y localicen manualmente lesiones en las retinografías. En este sentido, la principal contribución de este trabajo fue la aplicación de Explainable Artificial Intelligence como ayuda al diagnóstico de la DMAE. Para ello, desarrollamos un modelo de deep learning basado en RegNetY-320 con el que obtuvimos una precisión, sensibilidad y especificidad de 86.5%, 85.21% y 91.01%, respectivamente sobre la base de datos ADAM (1200 imágenes). Mediante la técnica integrated gradients, identificamos las áreas específicas en las retinografías que influyen en las decisiones del modelo. Esto permite la detección y localización de las lesiones asociadas a la DMAE, proporcionando una herramienta para la interpretación clínica y mejorando la confianza en el diagnóstico. Los resultados obtenidos indican que el método propuesto podría ser útil en los sistemas de ayuda al diagnóstico de la DMAE.

1. Introducción

La Degeneración Macular Asociada a la Edad (DMAE) se ha convertido en una causa principal de pérdida irreversible de la visión, especialmente en países desarrollados [1], [2]. Afecta principalmente a personas mayores de 60 años, y se prevé un aumento significativo de su prevalencia debido al envejecimiento demográfico, proyectándose más de 288 millones de casos para el año 2040 [2].

La DMAE se clasifica clínicamente en seca y húmeda [3]. La DMAE seca se caracteriza por la presencia de drusas de tamaño medio y cambios pigmentarios en la retina, mientras que la DMAE húmeda se caracteriza por la presencia de neovascularización y atrofia [3]. La retinografía y la tomografía de coherencia óptica son los exámenes auxiliares más utilizados en oftalmología. La fotografía del fondo de ojo es la prueba más económica y necesaria en la DMAE, ya que puede identificar de manera intuitiva lesiones y diagnosticar la enfermedad [3]. A pesar de su conveniencia, la implantación de programas de cribado de la DMAE a gran escala suele ser inviable, ya que la población de riesgo es grande y el análisis de las imágenes de fondo de ojo es muy complicado. Todo ello motiva la investigación de métodos de diagnóstico automáticos [4]–[8]. En la literatura, el enfoque predominante para el diagnóstico automatizado de la DMAE es entrenar un clasificador de aprendizaje automático para discriminar entre dos clases: DMAE y no DMAE. En algunos trabajos se han empleado métodos clásicos basados en redes neuronales totalmente conectadas [4] o máquinas de vectores de soporte [5]. Por el contrario, la mayoría de los trabajos recientes se basan en redes neuronales convolucionales (*Convolutional Neural Networks*, CNNs) [6]–[8]. Sin embargo, hasta donde sabemos, ninguno de ellos incorpora mecanismos de explicabilidad para ayudar a los expertos a comprender mejor las predicciones de los modelos. Este problema es particularmente relevante, ya que la ausencia de tales mecanismos limita la aplicación de enfoques automáticos en escenarios reales [9]. En tareas de ayuda al diagnóstico, esta necesidad de explicabilidad es aún más pronunciada, ya que la decisión del modelo puede tener un impacto directo en la salud del paciente [9].

Por ello, en este estudio, proponemos un enfoque de aprendizaje profundo que busca mejorar la detección automatizada de la DMAE y proporcionar explicaciones detalladas basadas en la identificación de las lesiones asociadas. El objetivo es desarrollar un sistema que combine la detección precisa de la DMAE en imágenes de retinografía con la capacidad de explicar las lesiones relacionadas, mejorando la precisión diagnóstica y la comprensión clínica de esta enfermedad ocular. Utilizando la técnica *integrated gradients*, generamos mapas de atribución que permitieron evaluar el impacto de cada píxel en la clasificación final de cada imagen y destacar las zonas más relevantes para la clasificación, contribuyendo así al campo de *Explainable Artificial Intelligence* (XAI).

2. Base de datos

En este estudio se utilizó la base de datos pública ADAM. Este conjunto de datos consiste en 1,200 imágenes de fondo de ojo almacenadas en formato JPEG, con 8 bits por canal de color. Estas imágenes fueron proporcionadas por el Centro Oftalmológico Zhongshan de la Universidad Sun Yat-sen en China [7]. Las imágenes de fondo de ojo se capturaron con un retinógrafo Zeiss Visucam 500, con una resolución de 2,124 × 2,056 píxeles, y con un retinógrafo Canon CR-2, con una resolución de 1,444 × 1,444 píxeles.

ISBN: 978-84-17853-76-1

El centro del campo de visión de las fotografías se ubicó en el disco óptico, la mácula o el punto medio entre el disco óptico y la mácula [7].

Todas las imágenes fueron etiquetadas indicando la presencia o no de DMAE. Las imágenes etiquetadas como DMAE abarcan casos de DMAE en sus etapas temprana, intermedia o avanzada, mientras que la categoría sin DMAE incluye muestras sin esta enfermedad pero que podrían presentar otros trastornos de retina [7].

Los creadores de la base de datos dividieron el conjunto de datos resultante en tres partes: un conjunto de entrenamiento (400 imágenes), un conjunto de validación (400 imágenes) y un conjunto de test (400 imágenes). En este estudio, se ha utilizado esta división.

3. Métodos

El método propuesto parte de una etapa de preprocesado para normalizar las imágenes de entrada. Después, se desarrolló un modelo CNN para detectar la DMAE haciendo uso de técnicas como *data augmentation*, *transfer learning* y *fine-tuning*. Finalmente, se utilizó la técnica i*ntegrated gradients* para visualizar las zonas de la imagen que influyen en las predicciones del modelo.

3.1. Preprocesado

Como entrada al modelo, se utilizaron versiones preprocesadas de las imágenes originales. En primer lugar, se redimensionaron todas las imágenes a 512 x 512 píxeles con el objetivo de normalizar todas las imágenes de entrada [10]. A continuación, los valores de los píxeles en cada canal de color de las imágenes se normalizaron en el intervalo [0,1]. Esto asegura que los valores de los píxeles estén en una escala consistente, facilitando el proceso de entrenamiento del modelo [11].

3.2. Data augmentation

Para aumentar el número de imágenes con las que entrenar la CNN, se empleó la técnica *data augmentation* en tiempo real. Esta técnica permite generar nuevas imágenes aleatorias exclusivas para cada época de entrenamiento. Para ello, se aplicaron trasformaciones simples sobre las 400 imágenes del conjunto de entrenamiento: rotaciones aleatorias en el rango [-90, +90] grados y volteos horizontales y verticales aleatorios [12].

3.3. *Transfer learning* y *fine-tuning*

El objetivo de aplicar *transfer learning* y *fine-tuning* es aprovechar el conocimiento adquirido por modelos previamente entrenados en tareas relacionadas. Esto permite acelerar el entrenamiento, mejorar el rendimiento y reducir los datos necesarios para lograr un modelo efectivo y preciso [13], [14].

La técnica *transfer learning* consiste en utilizar una red neuronal pre-entrenada para resolver un problema similar al que la red fue diseñada y entrenada inicialmente para resolver [13]. Para aplicar esta técnica, se utilizó un modelo base pre-entrenado con las imágenes del proyecto ImageNet, una base de datos que se compone de más de 14 millones de imágenes pertenecientes a más de 20,000 clases distintas [15].

A continuación, se aplicó la técnica *fine-tuning*. En los modelos basados en redes profundas, las primeras capas capturan características generales comunes a todas las imágenes, como bordes o patrones, mientras que las últimas capas identifican las características más relevantes para el problema a resolver [26]. La técnica de *fine-tuning* permite reentrenar todas o algunas de las capas finales de la CNN utilizando imágenes específicas del problema a resolver. De esta manera, el método resultante es más preciso en comparación con los modelos entrenados desde cero, ya que los pesos están mejor adaptados al objetivo del problema [14]. Si los datos son similares a los utilizados para pre-entrenar el modelo original, el *fine-tuning* de las últimas capas puede ser suficiente. Sin embargo, en este trabajo, se reentrenaron todas las capas del modelo ya que las imágenes tienen particularidades significativas. Reentrenar todas las capas permite una adaptación más efectiva, mejorando el rendimiento del modelo [14].

3.4. Arquitectura CNN

Las CNN son redes neuronales capaces de extraer características representativas de las imágenes de manera óptima y están formadas por distintos tipos de capas llamadas convolucionales, de *pooling* y *fully-connected*, entre otras [13]. En este trabajo se desarrolló una arquitectura CNN utilizando como modelo base la arquitectura RegNetY302 [16]. RegNet es un enfoque de diseño de redes que busca crear redes rápidas, simples y eficientes mediante una parametrización lineal. Estas redes superan a modelos populares como EfficientNet en rendimiento y velocidad en GPU. [16].

Para adaptar la arquitectura al problema de clasificación binaria de la DMAE, se han añadido una capa *average pooling*, una capa de *dropout* con un factor de 0.5, una capa densa de 2048 neuronas, otra capa de *dropout* con factor de 0.5 y una capa densa de 2 neuronas [17]. El número de neuronas de la última capa se corresponde con el número de clases que queremos discriminar: paciente sano y paciente con DMAE. En la primera capa densa se utilizó una función de activación tipo ReLU, mientras que en la última capa densa se utilizó la función de activación *softmax*. La función *softmax* devuelve una distribución de probabilidad sobre las clases. En este caso, la distribución de probabilidad se utiliza para indicar la probabilidad de que la imagen de entrada pertenezca a la clase sana o con DMAE [17]. En la fase de *fine-tuning*, se empleó la técnica *early-stopping* para minimizar el sobre-entrenamiento de la red [13]. De esta manera, el proceso de entrenamiento se detuvo automáticamente cuando la pérdida de validación no mejoraba durante 5 épocas. Se aplicó la entropía cruzada categórica como función de pérdida y Adam como algoritmo de optimización [17]. Asimismo, en épocas avanzadas, se redujo la tasa de aprendizaje en un factor de 10 cada vez que el error de validación alcanzase un mínimo y se mantuviese constante [13].

3.5. Integrated Gradients

Los modelos basados en *deep learning* son vistos como cajas negras que no explican cómo se realizan las predicciones. Esto ha llevado a una aceptación relativamente baja de estos modelos entre los profesionales

de la salud. Por lo tanto, se requiere una explicación e interpretación del funcionamiento del modelo en aplicaciones clínicas. En este sentido, en este trabajo proponemos un enfoque XAI que emplea el método de *integrated gradients* [18].

El método *integrated gradients* se basa en calcular la importancia de cada píxel en una imagen al sumar los gradientes de una serie de imágenes interpoladas entre una imagen de referencia y la imagen original [18]. La imagen de referencia representa la ausencia de las características que queremos analizar. Para elegir esta imagen de referencia nos basamos en estudios previos relacionados con el análisis de retinografías en retinopatía diabética (RD) puesto que, hasta donde tenemos conocimiento, no hay estudios que empleen este enfoque en el diagnóstico de DMAE. En los estudios de RD, es común utilizar una imagen negra como referencia [19]. En este estudio se utilizó la misma aproximación puesto que las imágenes de partida (retinografías) son iguales en ambos casos. Para cada retinografía, se generó una serie de 30 imágenes interpoladas entre la imagen de referencia negra y la imagen original. Estas imágenes son transiciones desde la imagen negra hasta la imagen de entrada original y representan diferentes niveles de importancia de las características [19].

Mediante *integrated gradientes* obtenemos mapas de atención que indican la contribución de cada píxel en la imagen original al resultado del diagnóstico. Estas contribuciones se miden en comparación con la imagen de referencia, que no aporta información relevante al modelo [18]. Esta técnica nos permite comprender cómo el modelo procesa la información y qué características de la imagen considera más relevantes durante el proceso de predicción.

4. Resultados

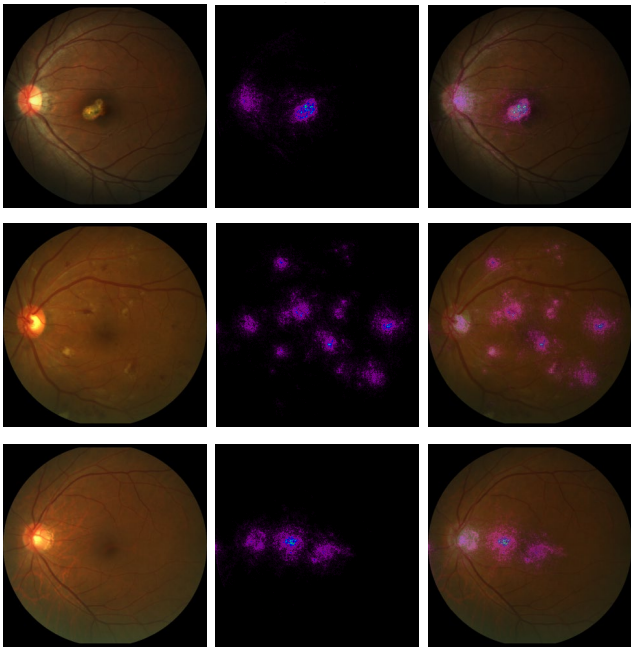
El entrenamiento de la CNN se llevó a cabo con 400 imágenes de la base de datos ADAM. Las 400 imágenes del conjunto de validación permitieron monitorizar el error de aprendizaje en la aplicación de la técnica *early-stopping*. El método se evaluó sobre un conjunto de test de 400 imágenes de la misma base de datos. La Tabla 1 recoge los resultados en términos de precisión (PR), sensibilidad (SE), especificidad (ES), F1-score y área bajo la curva ROC (AUC). Además, en la Figura 1 se presentan varios ejemplos de mapas obtenidos con el método *integrated gradients* donde se puede apreciar la identificación de lesiones en la clase con DMAE.

5. Discusión

En este estudio, se presenta una herramienta de explicabilidad basada en *deep learning* y el método *integrated gradients* para mejorar el diagnóstico de la de la DMAE. Es importante destacar que, hasta donde sabemos, este enfoque de atención no se ha empleado en el contexto del diagnóstico de la DMAE. Este método permite determinar si los resultados del modelo son fiables al resaltar las regiones importantes que conducen a la decisión del modelo. El método se ha desarrollado utilizando exclusivamente la base de datos ADAM. Los resultados de la Tabla 1 indican que el método propuesto ofrece un alto rendimiento para la detección de la DMAE. En la Tabla 2 se muestra que nuestros resultados son comparables a los obtenidos en estudios previos.

PR (%)	SE (%)	ES (%)	AUC
86.5	91.01	85.21	0.95

Tabla 2. *Resultados sobre el conjunto de test.*



(a) (b) (c)

Figura 1. *Ejemplos de mapas generados con integrated gradients: (**a**) imagenes originales, (**b**) mapas generados, (**c**) superposición de mapas generados e imágenes originales.*

La contribución más importante, en el contexto de XAI, son los mapas de atención generados con *integrated gradients*, que se validaron de forma cualitativa. Estos mapas desempeñan un papel fundamental al representar la relevancia de cada región de una imagen en el proceso de toma de decisiones del modelo. En la Figura 1 se presentan varios ejemplos de estos mapas sobre retinografías. En la primera fila, se puede apreciar que el modelo identifica de manera precisa las drusas cercanas al área de la mácula, lo cual constituye un indicativo relevante de la presencia de DMAE. La segunda fila representa un caso en el que una imagen fue incorrectamente clasificada como DMAE, posiblemente debido a otras lesiones. La última fila corresponde a un caso en el que una imagen se etiquetó incorrectamente como no DMAE, a pesar de la presencia de la enfermedad, lo que podría deberse a otros datos clínicos. El método *integrated gradients* proporciona a los

Estudio	Conjunto de test (*n*)	AUC
VUNO EYE TEAM [10]	ADAM (*n*=400)	0.97
ForbiddenFruit [10]	ADAM (*n*=400)	0.96
ADAM-TEAM [10]	ADAM (*n*=400)	0.93
XxlzT [10]	ADAM (*n*=400)	0.91
TeamTiger [10]	ADAM (*n*=400)	0.91
Airamatrix [10]	ADAM (*n*=400)	0.88
Método propuesto	ADAM (*n*=400)	0.95

Tabla 1. *Comparación de los resultados con otros estudios. n – Número de imágenes en el conjunto.*

ISBN: 978-84-17853-76-1

clínicos una mayor confianza en el diagnóstico al marcar claramente las lesiones asociadas a la enfermedad.

Este estudio presenta ciertas limitaciones que también es necesario mencionar. En primer lugar, el algoritmo ha sido desarrollado y validado únicamente en una base de datos. Sería deseable evaluar la capacidad de generalización de manera más exhaustiva, empleando un mayor número de imágenes, captadas con diferentes retinógrafos y procedentes de diversas localizaciones. Además, el método de visualización empleado representa una primera aproximación en términos de la interpretación y explicación de las predicciones del modelo. Aunque *integrated gradients* ofrece una visión de las áreas influyentes en las decisiones del modelo, existen otros mecanismos y enfoques de visualización basados en XAI que podrían ser explorados en investigaciones futuras. Abordar otros métodos podría ofrecer una comprensión más completa de cómo las CNNs toman decisiones.

6. Conclusiones

En este trabajo se ha propuesto un método automático basado en técnicas de *deep learning* para la detección de la DMAE que, además, incorpora un mecanismo de explicación visual para ayudar a los expertos a comprender mejor las predicciones del modelo. Los resultados obtenidos indican que método propuesto podría ser de utilidad en un entorno clínico como primera etapa en los sistemas de ayuda al diagnóstico de la DMAE.

Agradecimientos

Esta investigación se ha desarrollado en el marco de las ayudas TED2021-131913B-I00, PID2020-115468RB-I00 y PGC2018-098214-A-I00 financiadas por el 'Ministerio de Ciencia e Innovación/Agencia Estatal de Investigación/10.13039/501100011033/' y el Fondo Europeo de Desarrollo Regional (FEDER). Una forma de hacer Europa; y por el 'CIBER en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN)' a través del 'Instituto de Salud Carlos III' cofinanciado con fondos FEDER. M. Herrero Tudela cuenta con un contrato predoctoral de la Universidad de Valladolid.

Referencias

[1] C. J. Flaxel *et al.*, "Age-Related Macular Degeneration Preferred Practice Pattern®," *Ophthalmology*, vol. 127, no. 1, pp. P1–P65, Jan. 2020, doi: 10.1016/J.OPHTHA.2019.09.024.

[2] W. L. Wong *et al.*, "Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis," *Lancet Glob Health*, vol. 2, no. 2, pp. e106–e116, Feb. 2014, doi: 10.1016/S2214-109X(13)70145-1.

[3] L. S. Lim, P. Mitchell, J. M. Seddon, F. G. Holz, and T. Y. Wong, "Age-related macular degeneration," *The Lancet*, vol. 379, no. 9827, pp. 1728–1738, 2012.

[4] E. Pead *et al.*, "Automated detection of age-related macular degeneration in color fundus photography: a systematic review," *Surv Ophthalmol*, vol. 64, no. 4, pp. 498–511, Jul. 2019, doi: 10.1016/j.survophthal.2019.02.003.

[5] M. R. K. Mookiah *et al.*, "Automated detection of age-related macular degeneration using empirical mode decomposition," *Knowl Based Syst*, vol. 89, pp. 654–668, Nov. 2015, doi: 10.1016/J.KNOSYS.2015.09.012.

[6] C. González-Gonzalo *et al.*, "Evaluation of a deep learning system for the joint automated detection of diabetic retinopathy and age-related macular degeneration," *Acta Ophthalmol*, vol. 98, no. 4, pp. 368–377, Jun. 2020, doi: 10.1111/AOS.14306.

[7] H. Fang *et al.*, "ADAM Challenge: Detecting Age-related Macular Degeneration from Fundus Images," Feb. 2022, doi: 10.1109/TMI.2022.3172773.

[8] X. Li, M. Jia, M. T. Islam, L. Yu, and L. Xing, "Self-Supervised Feature Learning via Exploiting Multi-Modal Data for Retinal Disease Diagnosis," *IEEE Trans Med Imaging*, vol. 39, no. 12, pp. 4023–4033, Dec. 2020, doi: 10.1109/TMI.2020.3008871.

[9] B. H. M. van der Velden, H. J. Kuijf, K. G. A. Gilhuijs, and M. A. Viergever, "Explainable artificial intelligence (XAI) in deep learning-based medical image analysis," *Med Image Anal*, vol. 79, p. 102470, Jul. 2022, doi: 10.1016/J.MEDIA.2022.102470.

[10] H. Fang *et al.*, "ADAM Challenge: Detecting Age-Related Macular Degeneration From Fundus Images," *IEEE Trans Med Imaging*, vol. 41, no. 10, pp. 2828–2847, Oct. 2022, doi: 10.1109/TMI.2022.3172773.

[11] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift." 2015.

[12] L. Perez and J. Wang, "The Effectiveness of Data Augmentation in Image Classification using Deep Learning," Dec. 2017, Accessed: Jul. 12, 2023. [Online]. Available: https://arxiv.org/abs/1712.04621v1

[13] L. Shao, F. Zhu, and X. Li, "Transfer Learning for Visual Categorization: A Survey," *IEEE Trans Neural Netw Learn Syst*, vol. 26, no. 5, pp. 1019–1034, 2015, doi: 10.1109/TNNLS.2014.2330900.

[14] C. Käding, E. Rodner, A. Freytag, and J. Denzler, "Fine-Tuning Deep Neural Networks in Continuous Learning Scenarios," Jul. 2017, pp. 588–605. doi: 10.1007/978-3-319-54526-4_43.

[15] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255. doi: 10.1109/CVPR.2009.5206848.

[16] I. Radosavovic, R. P. Kosaraju, R. Girshick, K. He, and P. Dollár, "Designing Network Design Spaces," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 10425–10433, Mar. 2020, doi: 10.1109/CVPR42600.2020.01044.

[17] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[18] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic Attribution for Deep Networks," in *Proceedings of the 34th International Conference on Machine Learning*, D. Precup and Y. W. Teh, Eds., in Proceedings of Machine Learning Research, vol. 70. PMLR, Jul. 2017, pp. 3319–3328.

[19] R. Sayres *et al.*, "Using a Deep Learning Algorithm and Integrated Gradients Explanation to Assist Grading for Diabetic Retinopathy," *Ophthalmology*, vol. 126, no. 4, pp. 552–564, 2019, doi: https://doi.org/10.1016/j.ophtha.2018.11.016.