Aprendizaje semisupervisado usando retinografías en ayuda al diagnóstico de la degeneración macular asociada a la edad

R. Romero-Oraá[1,2], M. Herrero-Tudela[1], R. Hornero[1,2], M. I. López Gálvez[1,2], P. Romero-Aroca[3], M. García[1,2]

[1] Grupo de Ingeniería Biomédica, Universidad de Valladolid, Valladolid, España, {roberto.romero, maria.herrero.tudela, roberto.hornero, maria.garcia.gadanon}@uva.es

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

3 Servicio de oftalmología, Hospital Universitario Sant Joan de Reus, Institut d'Investigació Sanitària Pere Virgili [IISPV], Universitat Rovira i Virgili, Reus, Tarragona, España

Resumen

La degeneración macular asociada a la edad (DMAE) es una de las principales causas de discapacidad visual en las personas mayores. Los métodos de deep learning basados en aprendizaje supervisado son efectivos analizando retinografías para un diagnóstico precoz, pero requieren una gran cantidad de imágenes etiquetadas. Por el contrario, el aprendizaje no supervisado y, en particular, el aprendizaje autosupervisado (self-supervised learning, SSL) permiten explotar imágenes sin etiquetas. Si se combina SSL con una fase de ajuste fino utilizando un conjunto reducido de imágenes etiquetadas, hablamos de aprendizaje semisupervisado. En este trabajo se ha utilizado el modelo público RETFound, basado en SSL y previamente entrenado con 904170 retinografías. Como novedad en este trabajo, se llevó a cabo una etapa de ajuste fino empleando la base de datos ADAM, compuesta por 1200 imágenes. Esta es la primera vez que se utiliza este modelo para el diagnóstico de la DMAE. Los resultados alcanzaron una precisión del 89.50 %, una sensibilidad del 93.89 %, una especificidad del 74.16 %, un F1-score de 0.9329 y un área bajo la curva ROC de 0.9315. Estos resultados superan otras alternativas basadas en aprendizaje semisupervisado y están en línea con aquellos obtenidos aplicando transfer learning (aprendizaje supervisado). El trabajo desarrollado tiene potencial en un entorno clínico como método de ayuda al diagnóstico de la DMAE.

1. Introducción

La degeneración macular asociada a la edad (DMAE) es una de las principales causas de discapacidad visual grave en las personas mayores [1]. Afecta a 196 millones de personas en todo el mundo y se estima que esta cifra alcanzará los 288 millones en 2040 [1]. Un tratamiento adecuado requiere un diagnóstico precoz. Sin embargo, la enfermedad es asintomática en sus primeras etapas, por lo que son necesarios exámenes oftalmológicos periódicos del fondo de ojo [2]. Estos exámenes se basan comúnmente en el análisis de retinografías por ser la modalidad de imagen más rentable [3]. Dada la alta prevalencia de la DMAE y la falta de especialistas capacitados para su diagnóstico, los algoritmos automáticos de inteligencia artificial han demostrado ser útiles en el cribado de la enfermedad [3].

Los algoritmos más recientes de la literatura están basados en *deep learning*, que permite alcanzar altas precisiones sin la necesidad de extraer características de forma manual [2], [3]. Este tipo de algoritmos normalmente requieren una gran cantidad de imágenes etiquetadas en la fase de entrenamiento, puesto que se basan en aprendizaje supervisado [4]. Las retinografías son una modalidad de imagen común en la práctica clínica, con lo que es posible disponer de grandes volúmenes de imágenes [4]. No obstante, es complicado que estas imágenes estén etiquetadas, puesto que su interpretación clínica supone una elevada carga de trabajo para los especialistas [4]. En este sentido, el aprendizaje no supervisado y, en particular, el aprendizaje autosupervisado (*self-supervised learning*, SSL) son alternativas interesantes en este contexto, puesto que permiten explotar estas imágenes no etiquetadas para aprender representaciones de características generales que son comunes a todas ellas [4]. Posteriormente, es posible adaptar fácilmente ese aprendizaje a una tarea específica mediante una fase de ajuste fino (*fine-tuning*) utilizando un conjunto reducido de imágenes etiquetadas. Este procedimiento se conoce como aprendizaje semisupervisado [4].

Hasta donde sabemos, este tipo de aprendizaje solamente se ha utilizado con éxito para el diagnóstico de la DMAE en el trabajo de Li et al. [5]. Los autores combinan retinografías con angiofluoresceinografías generadas sintéticamente y proponen una función objetivo SSL que permite capturar la similitud entre pacientes y la invarianza entre transformaciones y entre ambas modalidades de imagen. No obstante, existen modelos SSL más avanzados y generalizables, como RETFound, que ha demostrado ser eficaz en el diagnóstico de la retinopatía diabética y el glaucoma [4]. Este modelo ha sido entrenado con casi un millón de retinografías y sus pesos son accesibles públicamente. Por ello, en este trabajo se propone aplicar aprendizaje semisupervisado utilizando el modelo RETFound con el objetivo de diagnosticar automáticamente la DMAE usando retinografías. Para ello, se partió del modelo original preentrenado con imágenes sin etiquetas (proporcionado por Zhout et al. [4]) y, como novedad, se aplicó a continuación una etapa de ajuste fino para adaptar el modelo a nuestro problema de clasificación binaria. Esta es la primera vez que se usa RETFound para el diagnóstico de la DMAE.

ISBN: 978-84-09-67332-2

2. Bases de datos de retinografías

Aunque la implementación y los pesos del modelo RETFound son accesibles públicamente, conviene detallar la manera en que fue desarrollado. Para entrenar el modelo RETFound aplicando SSL se utilizaron 904170 retinografías procedentes de las bases de datos Moorfields Diabetic imAge dataSet (MEH-MIDAS) [4] y Kaggle EyePACS [6]. MEH-MIDAS incluye 815468 retinografías de 37401 pacientes con diabetes que fueron atendidos en el hospital Moorfields Eye (Londres) entre el 2000 y el 2022 [4]. La base de datos Kaggle fue proporcionada por EyePACS en 2015 y está formada por 88702 retinografías capturadas en varios hospitales. Estas imágenes poseen distintos niveles de calidad y grados de severidad de la retinopatía diabética [6].

Para llevar a cabo la etapa de ajuste fino se utilizó la base de datos Automatic Detection challenge on Age-related Macular degeneration (ADAM), también conocida como Ichallenge-AMD, proporcionada por el Centro Oftalmológico Zhongshan de la Universidad de Sun Yat-sen (China) [3]. ADAM contiene 1200 retinografías etiquetadas según la presencia o ausencia de DMAE. Se respetó la división original de los grupos de entrenamiento, validación y test, todos ellos con 89 imágenes patológicas y 311 imágenes sanas (400 imágenes en total por grupo).

3. Métodos

En esta sección se describe, en primer lugar, el preprocesado aplicado a las imágenes de entrada. En segundo lugar, se detalla la construcción del modelo RETFound, tal y como lo proporcionan sus autores, incluyendo su arquitectura y estrategia de entrenamiento. Por último, se explica la fase de ajuste fino que permitió adaptar el modelo a nuestro problema.

3.1. Preprocesado

Como etapa de preprocesado, se eliminó el fondo negro que rodea el área retiniana y se aplicó aumentación de datos, incluyendo recortes aleatorios y volteos horizontales. Se redimensionaron las imágenes a 224 × 224 píxeles y se normalizaron en el rango [0-1] [4].

3.2. RETFound

Para construir la arquitectura de RETFound, se utilizó un autoencoder enmascarado (*masked autoencoder*, MAE), que está formado por un *encoder* y un *decoder*, tal y como se muestra en la Figura 1 [7]. Como *encoder* se empleó un *vision Transformer* con 24 bloques *transformer* y un vector de *embedding* de tamaño 1024. Como *decoder* se utilizó un *vision Transformer* con 8 bloques *transformer* y un vector de *embedding* de tamaño 512 [4].

Para llevar a cabo la estrategia SSL, la imagen se divide en parches de tamaño 16 × 16 y se enmascaran (eliminan) aleatoriamente el 75 % de ellos. El *encoder* se alimenta de los parches visibles (no enmascarados) y los proyecta en un vector de tamaño 1024 junto con su *embedding* posicional. A partir de este vector, los 24 bloques *transformer* generan un vector de características de alto nivel (*tokens*). A continuación, el *decoder* toma como entrada las características de alto nivel extraídas por el *encoder* añadiendo tokens ficticios para los parches enmascarados. Los 8 bloques *transformer* del decoder se encargan de reconstruir la imagen tras una proyección lineal. Una vez aplicada esta estrategia SSL, el *encoder* es capaz de generar características de alto nivel representativas de todas las retinografías [4].

Cabe destacar que los pesos del modelo RETFound entrenado con SSL son accesibles públicamente [4]. Por lo tanto, no fue necesario entrenar el modelo de cero. Para

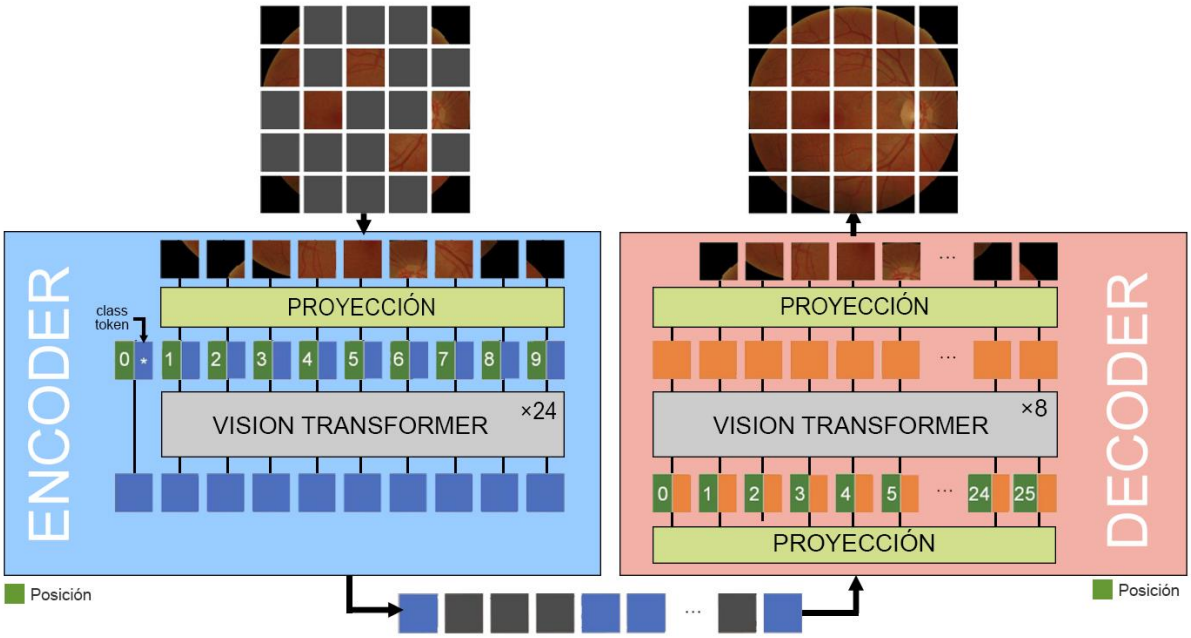

Figura 1. *Arquitectura de RETFound basada en un autoencoder enmascarado. El encoder toma parches visibles (no enmascarados) como entrada y genera características de alto nivel. El decoder recibe dichas características y añade características ficticias para los parches enmascarados. A su salida se reconstruyen los parches enmascarados (adaptado de Zhou et al. 2023 [4])*

ISBN: 978-84-09-67332-2

esta tarea, los autores emplearon 8 GPUs y un tamaño de *batch* de 1.792 imágenes. Entrenaron el modelo durante 800 épocas y aplicaron un calentamiento a la tasa de aprendizaje durante las primeras 15 (de 0 a $1 \times 10{-}3$).

3.3. Ajuste fino (*fine-tuning*)

Para construir la arquitectura empleada en esta etapa, se descartó el *decoder* y se utilizó solamente el *encoder* entrenado del modelo RETFound, tal y como se define en [4]. A su salida, se añadió un perceptrón multicapa compuesto por 3 capas *fully connected*. Este perceptrón toma como entrada las características de alto nivel extraídas por el *encoder* y calcula la probabilidad de pertenecer a cada clase (DMAE vs. No DMAE). La última de las capas *fully connected* estaba compuesta de 2 neuronas con función de activación *softmax*. La Figura 2 muestra la arquitectura empleada para esta fase.

Para llevar a cabo la etapa de ajuste fino se utilizó una GPU NVIDIA GeForce RTX 4080 con 16 GB de memoria. Se estableció un tamaño de *batch* de 12 imágenes y también se aplicó un calentamiento a la tasa de aprendizaje (en este caso, de 0 a $1 \times 10{-}5$). Como función de pérdidas se empleó la entropía cruzada y, como algoritmo de optimización, se utilizó AdamW [8]. Este algoritmo de descenso de gradiente estocástico se basa en la estimación adaptativa de los momentos de primer y segundo orden de la tasa de aprendizaje con una estrategia adicional de regularización [8]. Para el ajuste fino, el modelo se entrenó hasta un máximo de 200 épocas aplicando *early stopping* cuando el error sobre el conjunto de validación no mejoraba durante 7 épocas [9].

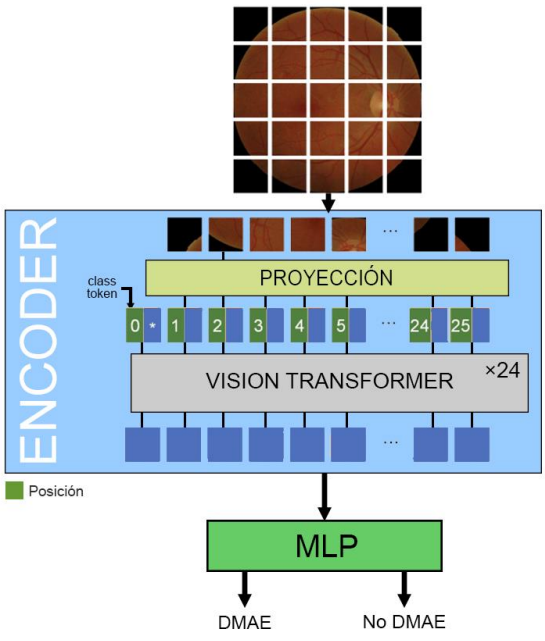

Figura 2. *Arquitectura empleada en la fase de ajuste fino. Se compone del encoder de RETFound y un perceptrón multicapa (adaptado de Zhou et al. 2023 [4]).*

4. Resultados

El modelo final fue evaluado sobre el conjunto de test de 400 imágenes de la base de datos ADAM. La Tabla 1 recoge los resultados en términos de precisión (PR), sensibilidad (SE), especificidad (ES), F1-score y área bajo la curva ROC (AUC).

5. Discusión

En este trabajo se ha explorado un método de aprendizaje semisupervisado para el diagnóstico automático de la DMAE. Para ello, se ha utilizado el modelo RETFound, basado en una arquitectura MAE y entrenado con casi un millón de imágenes aplicando SSL. Este modelo proporciona una solución generalizable para el procesado de retinografías que supera consistentemente a varios modelos anteriores en el diagnóstico de la retinopatía diabética y el glaucoma [4]. Sin embargo, hasta donde sabemos, RETFound nunca se había utilizado para el diagnóstico de la DMAE. Para adaptar el modelo a esta tarea, se utilizó una arquitectura compuesta por el *encoder* de RETFound y un perceptrón multicapa a la salida. Se llevó a cabo el ajuste fino utilizando 400 imágenes de entrenamiento y 400 de validación de la base de datos ADAM. Los resultados de la Tabla 1, obtenidos sobre las 400 imágenes de test, indican que la metodología empleada es muy efectiva para diagnosticar la DMAE contando con un reducido número de retinografías etiquetadas.

En la Tabla 1 también se muestran los resultados obtenidos por Xiaomeng et al. [5] sobre la base de datos ADAM. Estos autores también propusieron una estrategia SSL y alcanzaron una PR del 89.17 %, ligeramente inferior a la nuestra (89.50 %). No obstante, sus resultados son considerablemente inferiores a los nuestros para el resto de métricas (SE=83.17 % vs. SE=93.89 %, F1-score=0.8367 vs. F1-score=0.9329 y AUC=0.8317 vs. 0.9315), constatando que el modelo RETFound es más efectivo. En la comparación de la Tabla 1 también hemos incluido los resultados obtenidos en Romero-Oraá et al. [9], donde se exploró el uso de redes neuronales convolucionales (CNNs) con *transfer learning* (aprendizaje supervisado). Los resultados mostrados se corresponden con la arquitectura ResNet-RS preentrenada con ImageNet, que es la que alcanzó mayor rendimiento entre las arquitecturas comparadas. Los resultados alcanzados en [9] son muy similares a los obtenidos en este trabajo, con una PR del 89,50 % en ambos casos. En términos de AUC, la nueva propuesta se ve superada ligeramente por la anterior (AUC=0.9315 vs. 0.9497) y, en términos de ES, el valor alcanzado es notablemente inferior (74.16 % vs. 86.52 %). Sin embargo, nuestro valor de SE es mayor (93.89 % vs. 90.35 %), al igual que sucede para la métrica F1-score (0.9329 vs. 0.9305).

Otro aspecto destacable de la metodología es el poco tiempo que fue necesario para llevar a cabo la etapa de ajuste fino. Al partir de RETFound, un modelo preentrenado capaz de extraer características de alto nivel de las retinografías, la adaptación a nuestra tarea objetivo (ajuste fino) llevó tan solo 13 minutos con nuestro equipo.

ISBN: 978-84-09-67332-2

Estudio	PR (%)	SE (%)	ES (%)	F1-score	AUC
Xiaomeng et al. 2020	89.17	83.17	-	0.8367	0.8317
Romero-Oraá et al. 2023	**89.50**	90.35	**86.52**	0.9305	**0.9497**
Nuestro método	**89.50**	**93.89**	74.16	**0.9329**	0.9315

Tabla 1. *Comparación de los resultados obtenidos con distintos métodos para la base de datos ADAM.*

Aunque los resultados han sido satisfactorios, este estudio tiene algunas limitaciones. A nivel de rendimiento, se alcanzan resultados similares a los de la alternativa basada en *transfer learning*. No obstante, ambos métodos son complementarios y sería interesante desarrollar una estrategia combinada (*ensemble learning*). Otra limitación tiene que ver con la falta de interpretabilidad del modelo. En un futuro sería deseable aplicar técnicas de inteligencia artificial explicable para estudiar qué conocimiento aporta la fase SSL. Por último, cabe mencionar que sólo se utilizó una base de datos para el ajuste fino en nuestros experimentos. Sería conveniente replicar este estudio con bases de datos adicionales.

6. Conclusiones

En este trabajo se ha demostrado que el modelo RETFound es muy efectivo para analizar automáticamente retinografías en ayuda al diagnóstico de la DMAE. Los resultados alcanzados han superado otras alternativas basadas en aprendizaje semisupervisado y están en línea con aquellos obtenidos aplicando *transfer learning*. Por tanto, ambas estrategias tienen un gran potencial en el entorno clínico. Además, la combinación de ambas estrategias permitiría construir un método robusto de ayuda al diagnóstico de la DMAE.

Agradecimientos

Esta investigación se ha desarrollado en el marco de las ayudas TED2021-131913B-I00 y PID2020-115468RB-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] M. Fleckenstein, S. Schmitz-Valckenberg, and U. Chakravarthy, "Age-Related Macular Degeneration: A Review," *JAMA*, vol. 331, no. 2, pp. 147–157, Jan. 2024, doi: 10.1001/JAMA.2023.26074.

[2] S. Iqbal, T. M. Khan, K. Naveed, S. S. Naqvi, and S. J. Nawaz, "Recent trends and advances in fundus image analysis: A review," *Comput Biol Med*, vol. 151, p. 106277, Dec. 2022, doi: 10.1016/J.COMPBIOMED.2022.106277.

[3] 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.

[4] Y. Zhou *et al.*, "A foundation model for generalizable disease detection from retinal images," *Nature 2023 622:7981*, vol. 622, no. 7981, pp. 156–163, Sep. 2023, doi: 10.1038/s41586-023-06555-x.

[5] 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, 2020, doi: 10.1109/TMI.2020.3008871.

[6] Kaggle, "Diabetic Retinopathy Detection competition." [Online]. Available: https://www.kaggle.com/c/diabetic-retinopathy-detection/

[7] K. He, X. Chen, S. Xie, Y. Li, P. Dollar, and R. Girshick, "Masked Autoencoders Are Scalable Vision Learners," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2022-June, pp. 15979–15988, Nov. 2021, doi: 10.1109/CVPR52688.2022.01553.

[8] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," *7th International Conference on Learning Representations, ICLR 2019*, Nov. 2017, Accessed: Jul. 22, 2024. [Online]. Available: https://arxiv.org/abs/1711.05101v3

[9] R. Romero-Oraá, M. Herrero-Tudela, R. Hornero, M. I. López Gálvez, and M. García, "Comparación de múltiples redes neuronales convolucionales para el diagnóstico automático de la degeneración macular asociada a la edad usando retinografías," in *XLI Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2023)*, 2023, pp. 480–483.