

Comparación entre involuciones y convoluciones para la graduación de la severidad de la degeneración macular asociada a la edad

M. Herrero-Tudela¹, R. Romero-Oraá^{1,2}, R. Hornero Sánchez^{1,2}, Gonzalo C. Gutiérrez-Tobal^{1,2}, M. I. López Gálvez^{1,2}, P. Romero Aroca^{3,4}, 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

³ Servicio de Oftalmología, Hospital Universitario Sant Joan de Reus, Tarragona, España

⁴ Institut d'Investigació Sanitària Pere Virgili (IISPV), Universitat Rovira i Virgili, Tarragona, España

Resumen

La degeneración macular asociada a la edad (DMAE) es una de las principales causas de ceguera en personas mayores de 60 años en países desarrollados. Aunque existen tratamientos, su eficacia es limitada en etapas avanzadas, lo que hace crucial la detección temprana. Sin embargo, implementar programas de detección precoz es complicado debido a la extensa población de riesgo y a la necesidad de revisar manualmente las imágenes de fondo de ojo. El empleo de sistemas automáticos de ayuda al diagnóstico cobra relevancia en este contexto, siendo aquellos basados en redes neuronales convolucionales los predominantes en la última década. Sin embargo, en este estudio se pretende demostrar que la operación de involución también es efectiva para realizar este tipo de clasificación. Las involuciones generan núcleos de manera adaptativa basados en las características de entrada, promoviendo así la eficiencia. Para ello, se han comparado dos redes neuronales con el mismo número de capas: la primera basada en convoluciones, y la segunda en involuciones. La red neuronal convolucional alcanzó una precisión, sensibilidad y especificidad del 84.16 %, 68.33 % y 89.44 %, respectivamente, utilizando la base de datos pública AREDS. La red basada en involuciones superó ligeramente estos resultados, con una precisión del 85.08 %, sensibilidad de 70.17 % y especificidad de 90.05 %. Además, se han visualizado los kernels de involución para resaltar regiones específicas en las retinografías que influyen en las decisiones del modelo, proporcionando una herramienta para la interpretación clínica de la DMAE y mejorando la confianza en el diagnóstico.

1. Introducción

La degeneración macular asociada a la edad (DMAE) es una de las principales causas de pérdida visual en los países desarrollados, representando aproximadamente el 9 % de la ceguera mundial [1]. Su prevalencia aumenta exponencialmente con la edad, y se estima que en 2040 habrá 288 millones de personas con DMAE en el mundo [1, 2]. Esta enfermedad surge de una interacción compleja entre el envejecimiento, la genética y los factores de riesgo ambientales [3]. Se considera una enfermedad progresiva, y según el grado de severidad, se clasifica en tres etapas: temprana, intermedia y tardía [4]. Además, su detección temprana es crucial para prevenir una pérdida visual severa [4]. Con el envejecimiento de la población, el aumento en la incidencia de esta enfermedad podría

suponer una gran carga para los servicios oftalmológicos [2]. Detectar la DMAE en sus etapas iniciales puede ser complejo, ya que requiere revisiones oftalmológicas regulares y la captura de retinografías. En este contexto, los sistemas automáticos de análisis de imágenes adquieren particular relevancia en el diagnóstico de la enfermedad.

Diversos estudios han demostrado la eficacia de las herramientas automatizadas en el campo de la oftalmología [5, 6]. Los métodos basados en *deep learning* para la clasificación de la DMAE a partir de retinografías han mostrado buenos resultados [6–11]. Estos estudios han incluido sistemas binarios como el propuesto por Burlina et al. [7], donde usaron redes neuronales convolucionales (*Convolutional Neural Networks*, CNNs). El modelo que propusieron obtuvo precisiones de entre 88.4 % y 91.6 % en la detección de la presencia de la DMAE. Del mismo modo, se han empleado esquemas que combinan varias redes neuronales (*ensemble*) para distinguir entre 4 niveles de severidad (sin DMAE, DMAE temprana, DMAE intermedia y DMAE avanzada). En este trabajo se alcanzó una precisión del 95.3 % [9]. Otros autores han tratado de graduar la severidad de la enfermedad [8–10]. Por ejemplo, DeepSeeNet utiliza tres CNNs para clasificar las imágenes en 6 clases, con una precisión de 67.1 % y un coeficiente kappa de 0.558 sobre el conjunto de datos público AREDS [8]. Sobre la misma base de datos, Govindaiah et al [9] utilizaron una combinación de las redes Inception-ResNet-V2 y Xception, obteniendo una precisión del 86 % [9]. Por otro lado, Phan et al. [10] alcanzaron una precisión del 66.3 % con una máquina de vectores de soporte, también utilizando la base de datos AREDS.

Como se puede observar en los trabajos anteriores, la convolución sigue siendo el pilar fundamental de las redes neuronales profundas. La convolución se centra en la agnosticidad espacial, por lo que no permite capturar la correlación entre diferentes canales [12, 13]. Sin embargo, vale la pena estudiar cómo extraer tanto la información del dominio de los canales como la del dominio espacial. En este sentido, las características de la involución propuesta por Duo Li [12] permiten ayudar a descubrir la correlación entre canales para complementar la extracción de características básicas de la convolución [12]. En el presente trabajo, planteamos la

hipótesis de que debería ser posible construir un modelo basado en involuciones que tenga el mismo rendimiento o superior a uno basado en convoluciones para la graduación de la severidad de la DMAE en imágenes de fondo de ojo. Para validar esta hipótesis, se han comparado dos redes neuronales con una estructura de capas equivalente, una basada en convoluciones y otra en involuciones. Además, a diferencia de los trabajos anteriormente mencionados, el modelo propuesto basado en involuciones puede ofrecer información sobre el proceso de toma de decisiones, mostrando la salida de las involuciones y permitiendo visualizar las áreas de las imágenes que influyen en la predicción. De esta forma, el método proporcionaría una contribución natural también en el campo de la inteligencia artificial explicable (*Explainable Artificial Intelligence*, XAI) [14].

2. Base de datos

En este estudio, se utilizó el conjunto de datos público Age-Related Eye Disease Study (AREDS), perteneciente al National Institutes of Health de EE.UU. [15]. Esta base de datos fue diseñada para mejorar la comprensión de la progresión de la DMAE, su tratamiento y los factores de riesgo asociados [15]. Contiene retinografías capturadas con un campo de visión de 30° e incluye pares estereoscópicos de ambos ojos [15]. Expertos en oftalmología realizaron una evaluación del grado de severidad de la DMAE para todas las imágenes [15]. Para esta graduación, se empleó la escala de severidad simplificada AREDS, que se compone de 6 niveles: grado 1 (sin DMAE), grado 2 (DMAE temprana), grados 3a y 3b (DMAE intermedia) y grados 4a y 4b (DMAE tardía) [16]. Sin embargo, en un intento de maximizar la utilidad clínica, se determinó que diferenciar entre sí las dos etapas intermedias y los dos estadios tardíos carece de relevancia diagnóstica de cara al tratamiento. Por este motivo, en este estudio, los niveles intermedios (grados 3a y 3b) y tardíos (grados 4a y 4b) se agruparon, resultando en cuatro niveles de severidad [17].

El conjunto de datos público AREDS consta de más de 130,000 imágenes de fondo de ojo de 4,139 participantes de entre 55 y 80 años. Para el desarrollo de este trabajo se han empleado únicamente las imágenes con el campo de visión 2 (centrado en mácula) y se han excluido las imágenes de seguimiento para evitar redundancia. Esto hace un total de 10,366 imágenes, divididas en 80% entrenamiento (8,294 imágenes), 10% validación (1,036 imágenes) y 10% test (1,036 imágenes).

3. Métodos

En esta sección se describe, en primer lugar, el preprocesado aplicado a las imágenes de entrada. A continuación, se detallan las operaciones de convolución e involución y las arquitecturas a comparar, siendo la primera una CNN y la segunda una red neuronal involucional (*Involucional Neural Network*, INN).

3.1. Preprocesado

En primer lugar, las imágenes se recortaron alrededor del campo de visión, eliminando los bordes externos negros del fondo. A continuación, se redimensionaron todas las imágenes a 512×512 píxeles, con el objetivo de normalizar todas las imágenes de entrada y reducir el coste computacional. Finalmente, los valores de los píxeles en cada canal de color se normalizaron en el intervalo $[0,1]$ [6].

3.2. Aumento de datos

Las redes neuronales profundas requieren un gran número de datos de entrenamiento [6]. Para aumentar el número de imágenes con las que entrenar las redes de este trabajo, se empleó la técnica de aumento de datos en tiempo real, generando nuevas imágenes aleatorias exclusivas para cada época de entrenamiento [6]. Estas retinografías sintéticas se obtuvieron aplicando transformaciones simples como volteos verticales y horizontales aleatorios y ajustes aleatorios de contraste hasta un 10% [6].

3.3. Red neuronal convolucional

La convolución sigue siendo el pilar fundamental de las redes neuronales profundas en los métodos de procesado de imagen. Un esquema de la operación de convolución se muestra en la Figura 1. Al aplicar un conjunto de C_{out} núcleos de convolución de tamaño K, K, C_{in} sobre un tensor de entrada X con dimensiones H, W y C_{in} , se obtiene un tensor de salida Y con dimensiones H, W, C_{out} [12]. En la Figura 1, $C_{out} = 3$. Esto hace que el tensor de salida tenga una forma de H, W y 3. Se puede ver que el núcleo de convolución no depende de la posición espacial del tensor de entrada, lo que lo hace independiente de la ubicación. Por otro lado, cada canal en el tensor de salida se basa en un filtro de convolución específico, lo que lo hace dependiente del canal [12].

En este estudio se ha creado una red neuronal convolucional de tres capas convolucionales intercaladas con activaciones ReLU y MaxPooling. Al final de la red se han incluido dos capas densas de 64 y 4 neuronas para llevar a cabo la clasificación de las imágenes en los cuatro niveles de severidad.

3.4. Red neuronal involucional

Dadas las limitaciones inherentes a la operación de convolución, como su invarianza espacial y la redundancia entre

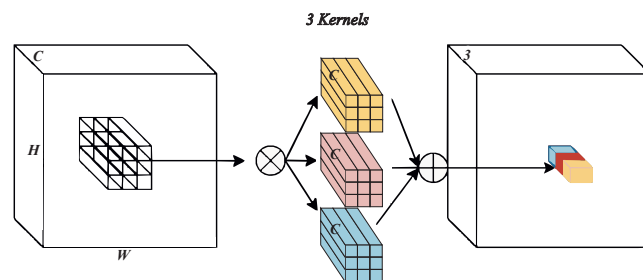


Figura 1: Esquema de la operación de convolución.

canales, surge la necesidad de desarrollar una operación que sea específica de la ubicación e independiente del canal. Con un número fijo de núcleos de involución (para cada posición espacial) no es posible procesar tensores de entrada de resolución variable [12]. Para abordar este problema, Duo Li et al. [12] propusieron la operación de involución como respuesta a la generación de cada núcleo condicionado a posiciones espaciales específicas. Este método permite procesar fácilmente tensores de entrada de resolución variable. La Figura 2 muestra un esquema de este método de generación de núcleos [12]. Dado un tensor de entrada X con dimensiones H , W y C_{in} , el tamaño de salida de la involución es el mismo que el tamaño de entrada, es decir, el tensor de salida tendrá dimensiones $H \times W \times C_{in}$. Por otra parte, los núcleos de involución se adaptan a cada píxel específico en la imagen, pero son compartidos a lo largo de los canales, es decir, tienen una dimensión espacial que se aplica a todas las canales de entrada C_{in} . Por lo tanto, el tamaño de los núcleos es $K \times K \times 1$, donde K es el tamaño espacial del núcleo. El mapa de características de salida de la involución se obtiene ejecutando operaciones de multiplicación y suma en la entrada con los núcleos de involución [12].

La arquitectura propuesta en este trabajo integra tres capas de involución con activaciones ReLU y MaxPooling para la reducción de la dimensión de las imágenes, finalizando con dos capas densas de 64 y 4 neuronas para llevar a cabo la clasificación.

4. Resultados

El modelo ha sido evaluado sobre un conjunto de test de 1,036 imágenes de la base de datos AREDS. En la Tabla 1 se recogen los resultados en términos de precisión (PR), sensibilidad (SE), especificidad (ES), y coeficiente kappa cuadrático, así como una comparación con estudios previos. Las matrices de confusión se ilustran en las Figuras 3a y 3b. Además, la salida de los núcleos involucionales se muestra en la Figura 4.

5. Discusión

En este trabajo se presenta un método basado en redes neuronales profundas para el estadiaje de la DMAE a partir de imágenes de fondo de ojo. Para ello se ha empleado una arquitectura novedosa basada en operaciones de involución. Se puede observar en las matrices de confusión (Figuras 3a y

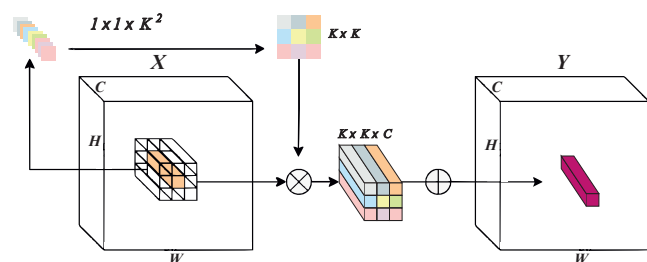
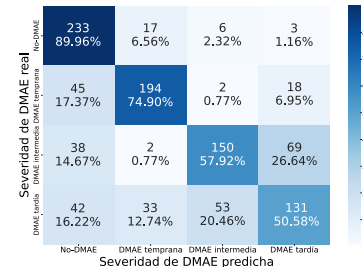


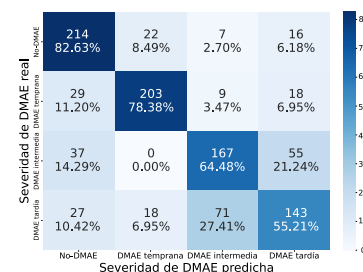
Figura 2: Esquema de la operación de involución (adaptado de Duo Li et al. (2021) [12])

Tabla 1: Comparación de los resultados obtenidos con distintos métodos en la detección de la DMAE.

Estudio	Base de datos	Número de imágenes	PR (%)	SE (%)	ES (%)	kappa
Phan et al (2016) [10]	Privada	279	62.7 %	-	-	57.3 %
Burlina et al (2017) [7]	AREDS	5,664	79.4 %	-	-	69.62 %
Y. Peng et al (2018) [8]	AREDS	59,302	67.1 %	-	-	55.8 %
Red neuronal convolucional propuesta	AREDS	10,366	84.16 %	68.33 %	89.44 %	64.12 %
Red neuronal involucional propuesta	AREDS	10,366	85.08 %	70.17 %	90.05 %	66.07 %



(a) Modelo convolucional



(b) Modelo involucional

Figura 3: Matrices de confusión.

3b) y las métricas correspondientes (Tabla 1), que los resultados para las dos arquitecturas propuestas son similares. No obstante, existen diferencias importantes entre ellas. En primer lugar, la CNN clásica tiene 67165508 parámetros, mientras que la INN tiene 786976 parámetros, teniendo el mismo número de capas en ambos casos. Por tanto, las redes basadas en involuciones presentan menos parámetros que las basadas en convoluciones, lo cual constituye una ventaja en términos de tiempo de computación. A la vista de los resultados en la Tabla 1, se puede observar como los resultados están en línea con estudios previos. Además, los estudios de Burlina et al. [7] y de Y. Peng et al. [8] hacen uso de las imágenes de ambos ojos para llevar a cabo la clasificación.

Una contribución importante en el contexto de XAI es la visualización de la salida de los núcleos de involución. Estos mapas representan la relevancia de cada región de una imagen en el proceso de toma de decisiones del modelo. En las imágenes de la Figura 4 se puede apreciar que el modelo identifica de manera precisa las lesiones cercanas al área de la mácula, un indicativo relevante de la presencia de DMAE.

Aunque los resultados han sido satisfactorios, este estudio tiene algunas limitaciones. En primer lugar, aunque la base de datos AREDS es, hasta donde tenemos conocimiento, la base de datos pública más amplia en el contexto de la DMAE, sería deseable evaluar la capacidad de generalización de manera más exhaustiva. Por otra parte, en futuros trabajos se tra-

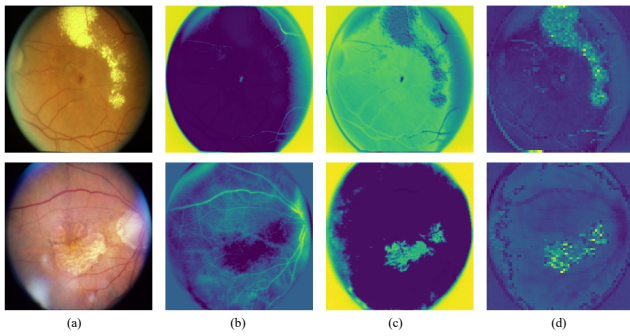


Figura 4: Visualización de la salida de los núcleos de involución. (a) Imagen original. (b) Salida del primer núcleo de involución. (c) Salida del segundo núcleo de involución. (d) Salida del tercer núcleo de involución.

tará de mejorar la tarea de clasificación a través de la combinación de las operaciones de involución y convolución, aprovechando al máximo las ventajas de ambas. Finalmente, es necesario mencionar que las arquitecturas propuestas en este trabajo toman únicamente una imagen de fondo de ojo como entrada. Puesto que AREDS dispone de la información de ambos ojos de los pacientes, en el futuro se tratará de ampliar el estudio para incluir una imagen de cada ojo del paciente en la graduación de la severidad de la DMAE.

6. Conclusiones

En este trabajo se ha demostrado que la operación de involución es efectiva en el análisis automático de retinografías para la ayuda al diagnóstico de la DMAE. Los resultados alcanzados han igualado otras alternativas basadas en CNNs. Por tanto, ambas estrategias tienen un gran potencial en el entorno clínico. El enfoque propuesto representa un avance significativo hacia el desarrollo de una herramienta diagnóstica confiable para la DMAE. La detección temprana de la DMAE puede facilitar un acceso oportuno al tratamiento y, en consecuencia, mejorar los resultados terapéuticos.

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.

El conjunto de datos utilizado para los análisis descritos en este trabajo fue obtenido de la Base de Datos Age-Related Eye Disease Study (AREDS) disponible en https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000001.v3.p1 a través del número de acceso dbGaP phs000001.v3.p1. El apoyo financiero

para AREDS fue proporcionado por National Eye Institute (N01-EY-0-2127). Queremos agradecer a los participantes de AREDS y al Grupo de Investigación de AREDS por su contribución a esta investigación.

Referencias

- [1] A. Quartilho, P. Simkiss, A. Zekite, W. Xing, R. Wormald, and C. Bunce, "Leading causes of certifiable visual loss in England and Wales during the year ending 31 March 2013," *Eye*, vol. 30, pp. 602–607, Apr 2016.
- [2] W. L. Wong, X. Su, X. Li, C. M. G. Cheung, R. Klein, C.-Y. Cheng, and T. Y. Wong, "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, pp. e106–116, Jan. 2014.
- [3] L. G. Fritsche, R. N. Fariss, D. Stambolian, G. R. Abecasis, C. A. Curcio, and A. Swaroop, "Age-related macular degeneration: genetics and biology coming together," *Annu Rev Genomics Hum Genet*, vol. 15, pp. 151–171, Apr. 2014.
- [4] F. L. Ferris, 3rd, C. P. Wilkinson, A. Bird, U. Chakravarthy, E. Chew, K. Csaky, S. R. Sadda, and Beckman Initiative for Macular Research Classification Committee, "Clinical classification of age-related macular degeneration," *Ophthalmology*, vol. 120, pp. 844–851, Jan. 2013.
- [5] R. Romero-Oraá, M. Herrero-Tudela, M. I. López, R. Hornero, and M. García, "Attention-based deep learning framework for automatic fundus image processing to aid in diabetic retinopathy grading," *Computer Methods and Programs in Biomedicine*, vol. 249, p. 108160, 2024.
- [6] M. Herrero-Tudela, R. Romero-Oraá, R. Hornero, G. C. Gutiérrez-Tobal, M. I. Lopez, and M. García, "An explainable deep-learning model to aid in the diagnosis of age related macular degeneration," in *9th European Medical and Biological Engineering Conference* (T. Jarm, R. Šmerc, and S. Mahnič-Kalamiza, eds.), (Cham), pp. 85–94, Springer Nature Switzerland, 2024.
- [7] P. M. Burlina, N. Joshi, M. Pekala, K. D. Pacheco, D. E. Freund, and N. M. Bressler, "Automated grading of Age-Related macular degeneration from color fundus images using deep convolutional neural networks," *JAMA Ophthalmol*, vol. 135, pp. 1170–1176, Nov. 2017.
- [8] Y. Peng, S. Dharssi, Q. Chen, T. D. Keenan, E. Agrón, W. T. Wong, E. Y. Chew, and Z. Lu, "DeepSeeNet: A deep learning model for automated classification of patient-based age-related macular degeneration severity from color fundus photographs," *Ophthalmology*, vol. 126, pp. 565–575, Nov. 2018.
- [9] A. Govindaiah, R. T. Smith, and A. Bhuiyan, "A new and improved method for automated screening of age-related macular degeneration using ensemble deep neural networks," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 702–705, 2018.
- [10] T. V. Phan, L. Seoud, H. Chakor, and F. Cheriet, "Automatic screening and grading of Age-Related macular degeneration from texture analysis of fundus images," *J Ophthalmol*, vol. 2016, p. 5893601, Apr. 2016.
- [11] P. Burlina, K. D. Pacheco, N. Joshi, D. E. Freund, and N. M. Bressler, "Comparing humans and deep learning performance for grading AMD: A study in using universal deep features and transfer learning for automated AMD analysis," *Comput Biol Med*, vol. 82, pp. 80–86, Jan. 2017.
- [12] D. Li, J. Hu, C. Wang, X. Li, Q. She, L. Zhu, T. Zhang, and Q. Chen, "Involution: Inverting the inheritance of convolution for visual recognition," 2021.
- [13] M. Jaderberg, A. Vedaldi, and A. Zisserman, "Speeding up convolutional neural networks with low rank expansions," 2014.
- [14] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bannetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," *Information Fusion*, vol. 58, pp. 82–115, 2020.
- [15] "The age-related eye disease study system for classifying age-related macular degeneration from stereoscopic color fundus photographs: the age-related eye disease study report number 6," *American Journal of Ophthalmology*, vol. 132, no. 5, pp. 668–681, 2001.
- [16] F. L. Ferris, M. D. Davis, T. E. Clemons, L.-Y. Lee, E. Y. Chew, A. S. Lindblad, R. C. Milton, S. B. Bressler, R. Klein, and Age-Related Eye Disease Study (AREDS) Research Group, "A simplified severity scale for age-related macular degeneration: AREDS report no. 18," *Arch Ophthalmol*, vol. 123, pp. 1570–1574, Nov. 2005.
- [17] A. Bhuiyan, T. Y. Wong, D. S. W. Ting, A. Govindaiah, E. H. Souied, and R. T. Smith, "Artificial Intelligence to Stratify Severity of Age-Related Macular Degeneration (AMD) and Predict Risk of Progression to Late AMD," *Translational Vision Science & Technology*, vol. 9, pp. 25–25, 04 2020.