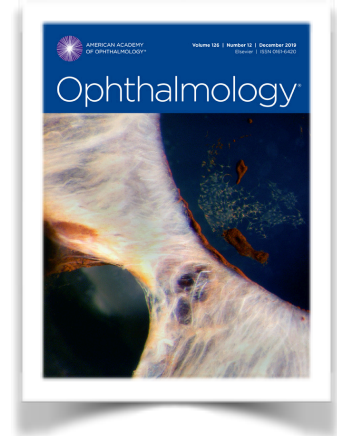


### **Deep Learning and Glaucoma Specialists** **The Relative Importance of Optic Disc Features to Predict** **Glaucoma Referral in Fundus Photographs**

**Sonia Phene, R. Carter Dunn, Naama Hammel, Yun Liu, et al.**



### **COMENTARIOS**

Necesitamos una herramienta para detección de glaucoma puesto que afecta a unos 70 millones de personas y entre el 50-90% lo desconocen. Para posibilitar la detección eficaz de glaucoma en entornos donde no hay médicos capacitados para evaluar la papila (cabeza del nervio óptico: ONH), este estudio desarrolla y valida un algoritmo de aprendizaje profundo ó *Deep Learning* (DL). Este algoritmo de DL a partir de retinografías puede detectar la presencia de características patológicas de la papila y Capa de Fibras Nerviosas de la Retina (RNFL) y así obtener información sobre qué características de ONH y RNFL impulsan a los especialistas de glaucoma a realizar la valoración de glaucoma. Pero a diferencia por ejemplo del despistaje de retinopatía diabética, el diagnóstico de glaucoma es mas complejo, porque también precisa datos clínicos, pruebas funcionales y estructurales, lo que puede limitar el valor del diagnóstico de los algoritmos de aprendizaje únicamente desarrollados únicamente a partir de imágenes de fondo de ojo.

Se desarrolló y validó un algoritmo de aprendizaje con 86618 imágenes obtenidas de manera retrospectiva de 3 bases de datos independientes y de diferentes poblaciones.

- Base A de EyePACS con 1205 imágenes, (resultando 18,1% referible para estudio de glaucoma).
- Conjunto de datos B, de un programa de detección teleretiniana diabética (9642 imágenes con 9,2% referible).
- Conjunto de datos C con imágenes de una clínica de glaucoma (346 imágenes, 81,7% referible).

Y se abordaron limitaciones de trabajos anteriores como:

1. Las imágenes de este estudio contaron con la valoración de un especialista en glaucoma lo que mejora la validez de la identificación de glaucoma.
2. Se definió el término glaucoma para el estándar de referencia, para aumentar la efectividad de evaluación del algoritmo.
3. No se excluyeron imágenes de baja calidad, para valorar la utilidad del algoritmo en entornos de la vida real.
4. No se excluyeron imágenes en las que los calificadores no estaban de acuerdo con el conjunto de validación, para evitar sesgar el conjunto de datos hacia casos más fáciles.
5. Se utilizaron imágenes completas de fondo de ojo en lugar de ampliadas en el nervio óptico, para poder obtener la evaluación de la RNFL.

Con todo ello, se comparó el rendimiento del algoritmo desarrollado (con las imágenes de estas 3 bases de datos tomadas como correctas) con la opinión de 43 evaluadores (entre ellos glaucomatólogos, oftalmólogos generales y optometristas) para comparar la concordancia diagnóstica y determinar la importancia relativa de esas características fundoscópicas en la decisión de derivación del paciente para estudio completo. Se definió glaucoma referible como: apariencia de ONH lo suficientemente preocupante como para justificar la derivación para un examen exhaustivo: Excavación  $>0.7$ , existencia de muescas, de defectos RNFL, vasos circumlineares, poros en la lámina cribosa, hemorragias papilares, nasalización de los vasos, ANR inferior  $<$  Superior y ANR Superior  $<$  Temporal y existencia de atrofia peripapilar beta.

## Resultados

El algoritmo logró un área bajo la curva (AUC) para referir glaucoma de 0.945 (intervalo de confianza [IC] del 95%, 0.929-0.960) en el conjunto de datos A, 0.855 (IC del 95%, 0.841-0.870) en el conjunto de datos B y 0.881 (IC del 95%, 0.838-0.918) en el conjunto de datos C. Las AUC de algoritmo oscilaron entre 0.661 y 0.973 para las características de ONH glaucomatosas.

Los 43 evaluadores analizaron las imágenes para detectar características glaucomatosas en la ONH y además se analizó la importancia relativa que los especialistas le atribuyeron a las diferentes características de la retinografía y si el algoritmo de aprendizaje aprendió una relación similar. Tanto para los glaucomatólogos como para el algoritmo, las características de la papila que se correlacionaron con mayor frecuencia con el diagnóstico de glaucoma fueron: Excavación Anillo Neuroretiniano  $>0.7$ , existencia de muescas, defectos de capa de fibras y vasos desnudos circumlineares.

Este algoritmo, mantuvo una buena correlación con los resultados de un estudio completo de glaucoma. Así demuestran, que un algoritmo de "Deep

Learning” entrenado solo en fotografías de fondo de ojo que contienen la papila, puede predecir con precisión las evaluaciones de glaucoma realizadas por especialistas en glaucoma con acceso a información adicional del paciente. El algoritmo mostró una sensibilidad significativamente mayor que 7 de 10 calificadores que no participaron en la determinación del estándar de referencia, incluyendo 2 de 3 glaucomatólogos, y mostró mayor especificidad que 3 calificadores (incluyendo 1 glaucomatólogo), sin dejar de ser comparable a otros.

El algoritmo puede ayudar a mejorar la decisión clínica de derivar pacientes a un especialista de glaucoma con una retinografía. Esta herramienta presenta una oportunidad para que el diagnóstico de glaucoma pueda llegar a poblaciones desatendidas antes de que ocurra una pérdida permanente de visión.

### **Ophthalmology. 2019; 126: 1627-39.**

Comentario realizado por la **Dra. María Rojo Arnao**. Hospital Virgen de la Luz (Cuenca).

## **ABSTRACT**

**PURPOSE** To develop and validate a deep learning (DL) algorithm that predicts referable glaucomatous optic neuropathy (GON) and optic nerve head (ONH) features from color fundus images, to determine the relative importance of these features in referral decisions by glaucoma specialists (GSs) and the algorithm, and to compare the performance of the algorithm with eye care providers.

**DESIGN** Development and validation of an algorithm.

**PARTICIPANTS** Fundus images from screening programs, studies, and a glaucoma clinic.

**METHODS** A DL algorithm was trained using a retrospective dataset of 86 618 images, assessed for glaucomatous ONH features and referable GON (defined as ONH appearance worrisome enough to justify referral for comprehensive examination) by 43 graders. The algorithm was validated using 3 datasets: dataset A (1205 images, 1 image/patient; 18.1% referable), images adjudicated by panels of GSs; dataset B (9642 images, 1 image/patient; 9.2% referable), images from a diabetic teleretinal screening program; and dataset C (346 images, 1 image/patient; 81.7% referable), images from a glaucoma clinic.

**MAIN OUTCOME MEASURES** The algorithm was evaluated using the area under the receiver operating characteristic curve (AUC), sensitivity, and specificity for referable GON and glaucomatous ONH features.

**RESULTS** The algorithm’s AUC for referable GON was 0.945 (95% confidence interval [CI], 0.929e0.960) in dataset A, 0.855 (95% CI, 0.841e0.870) in dataset B, and 0.881 (95% CI, 0.838e0.918) in dataset C. Algorithm AUCs ranged between 0.661 and 0.973 for glaucomatous ONH features. The algorithm showed significantly higher sensitivity than 7 of 10 graders not involved in determining the reference standard, including 2 of 3 GSs, and showed higher specificity than 3 graders (including 1 GS), while remaining comparable to others. For both GSs and the algorithm, the most crucial features related to referable GON were: presence of vertical cup-to-disc ratio of 0.7 or more, neuroretinal rim notching, retinal nerve fiber layer defect, and bared circumlinear vessels.

**CONCLUSIONS** A DL algorithm trained on fundus images alone can detect referable GON with higher sensitivity than and comparable specificity to eye care providers. The algorithm maintained good performance on an independent dataset with diagnoses based on a full glaucoma workup.