

ARTÍCULO DESTACADO DEL MES



Evaluation of a Deep Learning System For Identifying Glaucomatous Optic Neuropathy Based on Color Fundus Photographs

Lama A. Al-Aswad, Rahul Kapoor, Chia Kai Chu, Stephen Walters, et al.



COMENTARIOS

Presentamos la revisión de este artículo cuyo propósito es evaluar el desarrollo de un sistema de Inteligencia Artificial (IA) basado en el “deep learning” para la correcta identificación de la neuropatía óptica glaucomatosa.

Para ello se comparan los resultados del sistema **Pegasus** (Visulytix Ltd, London UK), frente a los resultados de 6 oftalmólogos en diferentes grados de formación (dos fellows en glaucoma, dos especialistas en glaucoma y dos especialistas en neuro-oftalmología) en su capacidad para orientar el diagnóstico glaucoma-no glaucoma (del 1 al 7) mediante el análisis de retinografías en color. Fueron analizadas 110 fotografías obtenidas de forma aleatoria de la base de datos **ORIGA** (Online Retinal fundus Image database for Glaucoma Analysis) que se creó para el Singapore Malay Eye Study (SiMES). Las imágenes fueron introducidas a la base de datos/nube del sistema Pegasus para ser analizadas en formato JPEG 3072*2048 p.

Ambos resultados, tanto humanos como de Pegasus, fueron comparados con el diagnóstico clínico original (SiMES), que fue definido como el “gold standard”, y fueron evaluados en términos de sensibilidad, eficacia y área bajo la curva (ROC). Además el rendimiento del sistema de aprendizaje se comparó con el escenario de consenso del “mejor caso”, que se determinó tomando el promedio de los resultados de los 5 mejores oftalmólogos.

Pegasus logró un área bajo la curva del 92.6% en comparación con los oftalmólogos que oscilaron entre 69.6% y 84.9% y el escenario de consenso de "mejor caso" del 89.1%. Pegasus tuvo una sensibilidad del 83.7% y una especificidad del 88.2%, mientras que la sensibilidad de los oftalmólogos varió del 61.3% al 81.6% y la especificidad varió del 80.0% al 94.1%. El acuerdo entre Pegasus y el gold standard fue de 0.715, mientras que el acuerdo más alto del oftalmólogo con el gold standard fue del 0.613. El acuerdo intra observador varió de 0.62 a 0.97 para los oftalmólogos y fue perfecto (1.00) para Pegasus. El sistema deep learning tardó aproximadamente el 10% del tiempo de los oftalmólogos para determinar la clasificación.

Entre las **fortalezas** de este estudio están que fueron incluidos evaluadores con diversos grados de experiencia y que fueron enmascarados con los resultados del análisis del sistema de aprendizaje profundo, la distribución de controles normales/pacientes con glaucoma, y con imágenes repetidas utilizadas para evaluar el acuerdo intra observador.

Las **limitaciones** de este estudio incluyen la estandarización del conjunto de datos, con la adquisición de la misma cámara y campo de visión del fondo de ojo, así como el tamaño de muestra relativamente pequeño. El tamaño de la muestra dificultó la realización de estudios informativos sobre el rendimiento del sistema de aprendizaje profundo en diferentes subconjuntos del conjunto de datos, por ejemplo: ¿es diferente el rendimiento del sistema de aprendizaje profundo cuando solo se utilizan ejemplos de atrofia peripapilar?. En el futuro se podrá estudiar el rendimiento del sistema en imágenes con diferentes características, como la atrofia peri papilar. Como conclusión se puede comentar que Pegasus superó a 5 de los 6 oftalmólogos en términos de rendimiento diagnóstico, y no hubo diferencias estadísticamente significativas entre el sistema de aprendizaje profundo y el consenso del "mejor caso" entre los oftalmólogos. La alta sensibilidad de Pegasus lo convierte en una herramienta valiosa para la detección de pacientes con neuropatía óptica glaucomatosa. En un futuro próximo estas herramientas serán incorporadas a los diferentes instrumentos diagnósticos utilizados en nuestra práctica clínica (CV, OCTs, retinógrafos) y nos ayudarán a realizar un diagnóstico más precoz, mejorando así el pronóstico de esta patología.

J Glaucoma 2019 Dec;28(12):1029-1034. doi: 10.1097/IJG.0000000000001319.

Comentario realizado por el **Dr. Aitor Fernández**. IOA Madrid Innova Ocular.

ABSTRACT

PRECIS:

Pegasus outperformed 5 of the 6 ophthalmologists in terms of diagnostic performance, and there was no statistically significant difference between the deep learning system and the "best case" consensus between the ophthalmologists. The agreement between Pegasus and gold standard was 0.715, whereas the highest ophthalmologist agreement with the gold standard was 0.613. Furthermore, the high sensitivity of Pegasus makes it a valuable tool for screening patients with glaucomatous optic neuropathy.

PURPOSE:

The purpose of this study was to evaluate the performance of a deep learning system for the identification of glaucomatous optic neuropathy.

MATERIALS AND METHODS:

Six ophthalmologists and the deep learning system, Pegasus, graded 110 color fundus photographs in this retrospective single-center study. Patient images were randomly sampled from the Singapore Malay Eye Study. Ophthalmologists and Pegasus were compared with each other and to the original clinical diagnosis given by the Singapore Malay Eye Study, which was defined as the gold standard. Pegasus' performance was compared with the "best case" consensus scenario, which was the combination of ophthalmologists whose consensus opinion most closely matched the gold standard. The performance of the ophthalmologists and Pegasus, at the binary classification of nonglaucoma versus glaucoma from fundus photographs, was assessed in terms of sensitivity, specificity and the area under the receiver operating characteristic curve (AUROC), and the intraobserver and interobserver agreements were determined.

RESULTS:

Pegasus achieved an AUROC of 92.6% compared with ophthalmologist AUROCs that ranged from 69.6% to 84.9% and the "best case" consensus scenario AUROC of 89.1%. Pegasus had a sensitivity of 83.7% and a specificity of 88.2%, whereas the ophthalmologists' sensitivity ranged from 61.3% to 81.6% and specificity ranged from 80.0% to 94.1%. The agreement between Pegasus and gold standard was 0.715, whereas the highest ophthalmologist agreement with the gold standard was 0.613. Intraobserver agreement ranged from 0.62 to 0.97 for ophthalmologists and was perfect (1.00) for Pegasus. The deep learning system took ~10% of the time of the ophthalmologists in determining classification.

CONCLUSION:

Pegasus outperformed 5 of the 6 ophthalmologists in terms of diagnostic performance, and there was no statistically significant difference between the deep learning system and the "best case" consensus between the ophthalmologists. The high sensitivity of Pegasus makes it a valuable tool for screening patients with glaucomatous optic neuropathy. Future work will extend this study to a larger sample of patients.