



Performances of Machine Learning in Detecting Glaucoma Using Fundus and Retinal Optical Coherence Tomography Images: A Meta-Analysis

Jo-Hsuan Wu, Takashi Nishida, Robert N Weinreb, Jou-Wei Lin



COMENTARIOS

En glaucoma, de todos es conocido, es fundamental la detección y un cribado preciso de la enfermedad. La evaluación del glaucoma por parte de los médicos especialistas requiere tiempo, trabajo y habilidades específicas que pueden conducir a la inconsistencia y la subjetividad, por todo ello la aplicación de imagen automatizada con el análisis de Machine Learning (ML) se postula como una herramienta más en la mejora en su detección.

ML usa datos existentes para entrenar y construir un modelo que realiza automáticamente ciertas tareas en un nuevo conjunto de datos, proporcionando un enfoque menos dependiente del ser humano para la interpretación de imágenes.

Además de los algoritmos convencionales de ML como support vector machine (SVM); random forest o K-nearest neighbor, existen nuevas técnicas como la neuronal network o red neuronal artificial entrenada (NN) que han transformado el análisis de imagen automatizada, demostrando por ejemplo mayor precisión en la detección de retinopatía diabética en comparación con expertos especialistas. Aunque muchos estudios han mostrado resultados positivos en la clasificación de glaucoma, la evidencia de su desempeño general aún es limitada.

El trabajo que analizamos pretende mediante un metanálisis evaluar el rendimiento de *Machine Learning* (ML) en la detección de glaucoma usando retinografías e imágenes de OCT. Para ello realiza búsquedas en PubMed y EMBASE publicados hasta el 11 de agosto de 2021. Se emplea modelo bivariados de efectos aleatorios (*bivariate random-effects model*) para evaluar: sensibilidad, especificidad, área bajo la curva (AUC) y cocientes de probabilidad positivos y negativos.

La búsqueda arrojó 1565 y 1622 registros de PubMed y EMBASE, respectivamente. Finalmente, se incluyeron 105 estudios para el análisis de datos. El número total de pruebas fueron de 197.174 para retinografías y 16.039 para OCT. En general, ML mostró excelentes resultados para retinografías (sensibilidad 0,92 [IC 95%, 0,91-0,93]; especificidad 0,93 [IC del 95%, 0,91-0,94] y AUC 0,97 [IC 95%, 0,95-0,98]) y OCT (sensibilidad 0,90 [IC 95%, 0,86-0,92]; especificidad 0,91 [IC 95%, 0,89-0,92] y AUC 0,96 [IC 95%, 0,93-0,97]).

Dentro de los diferentes clasificadores de ML, para las retinografías los mejores resultados se obtuvieron con SVM con una media agrupada de sensibilidad de 0,96 (IC 95%, 0,93-0,97), especificidad de 0,97 (IC 95%, 0,94-0,98), y AUROC de 0,99 (IC 95% 0,98-1,00). NN y otro tuvieron resultados ligeramente inferiores pero similares.

Para OCT, aunque la AUROC fue similar para las 3 categorías, SVM mostró una mayor sensibilidad (0,92; 0,95% IC, 0,84-0,96) y especificidad (0,95; IC 95%, 0,93-0,97), en comparación con NN y otros.

Conclusión:

El rendimiento de ML mostró un excelente rendimiento diagnóstico utilizando ambos tipos de imágenes, como lo demuestra la alta sensibilidad, especificidad y AUROC. Comparando diferentes métodos de clasificación de ML, SVM mostró rendimiento más alto entre las 3 categorías. Aunque NN y otro fueron ligeramente inferiores, la precisión diagnóstica seguía siendo buena.

La precisión de ML en la detección de glaucoma es probablemente comparable con la de los expertos, como lo demuestra el alto AUROC general de más de >95% para retinografías y OCT.

Los resultados satisfactorios indican que la realización de ML puede ser útil en la práctica diaria. Aún así, se deben atender algunos factores potenciales que afectan al desempeño de ML, particularmente **miopía magna**, que se informó como responsable de aproximadamente el 50% de los falsos negativos y positivos por medio de ML.

Existen limitaciones a tener en cuenta:

-El glaucoma presenta diferentes subtipos y se desconoce si los resultados pueden ser aplicados a un tipo o etapa en concreto de la enfermedad.

-Existe una alta heterogeneidad al realizar la evaluación, esperable debido a la gran cantidad de estudios incluidos en el diseño de ML.

-En algunos estudios se apreció una falta de transparencia, aunque no debería afectar al análisis actual.

- Estudios prospectivos son necesarios para comparar directamente el rendimiento de ML con los expertos.

En conclusión, el rendimiento de la técnica ML en la detección de glaucoma utilizando retinografías y OCT se compara favorablemente con la de los expertos, y su excelente precisión es prometedora para la aplicación clínica.

Am J Ophthalmol . 2022 May;237:1-12. doi: 10.1016/j.ajo.2021.12.008

Comentario realizado por la **Dra. Teresa Laborda. Hospital La Arruzafa. Córdoba.**

ABSTRACT

Purpose: To evaluate the performance of machine learning (ML) in detecting glaucoma using fundus and retinal optical coherence tomography (OCT) images.

Design: Meta-analysis.

Methods: PubMed and EMBASE were searched on August 11, 2021. A bivariate random-effects model was used to pool ML's diagnostic sensitivity, specificity, and area under the curve (AUC). Subgroup analyses were performed based on ML classifier categories and dataset types.

Results: One hundred and five studies (3.3%) were retrieved. Seventy-three (69.5%), 30 (28.6%), and 2 (1.9%) studies tested ML using fundus, OCT, and both image types, respectively. Total testing data numbers were 197,174 for fundus and 16,039 for OCT. Overall, ML showed excellent performances for both fundus (pooled sensitivity = 0.92 [95% CI, 0.91-0.93]; specificity = 0.93 [95% CI, 0.91-0.94]; and AUC = 0.97 [95% CI, 0.95-0.98]) and OCT (pooled sensitivity = 0.90 [95% CI, 0.86-0.92]; specificity = 0.91 [95% CI, 0.89-0.92]; and AUC = 0.96 [95% CI, 0.93-0.97]). ML performed similarly using all data and external data for fundus and the external test result of OCT was less robust (AUC = 0.87). When comparing different classifier categories, although support vector machine showed the highest performance (pooled sensitivity, specificity, and AUC ranges, 0.92-0.96, 0.95-0.97, and 0.96-0.99, respectively), results by neural network and others were still good (pooled sensitivity, specificity, and AUC ranges, 0.88-0.93, 0.90-0.93, 0.95-0.97, respectively). When analyzed based on dataset types, ML demonstrated consistent performances on clinical datasets (fundus AUC = 0.98 [95% CI, 0.97-0.99] and OCT AUC = 0.95 [95% 0.93-0.97]).

Conclusions: Performance of ML in detecting glaucoma compares favorably to that of experts and is promising for clinical application. Future prospective studies are needed to better evaluate its real-world utility.