Title: AI For Glaucoma Diagnosis: A Scoping Review and Evidence Map

Authors: Samantha Braverman[1], Louis Leslie[2], Su-Hsun Liu[2,3], Tianjing Li[2,3]

Affiliations: [1]School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA [2]Department of Ophthalmology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA [3]Department of Epidemiology, Colorado School of Public Health, Aurora, CO, USASupport: This is project is support by the National Eye Institute, National Institutes of Health (grant #: UG1 EY020522).

Disclosures: All authors report no commercial relationship

Purpose:
Screening for glaucoma is traditionally time-consuming and resource-intensive. Artificial Intelligence (AI) has been proposed as a means to reduce this burden and broaden screening efforts. However, the extent of AI's implementation within the field remains unclear. This study aims to assess the span of literature on diagnostic test accuracy (DTA) studies comparing AI with human graders for glaucoma diagnosis.

Methods:
On September 3, 2024, a search of Cochrane Center Register of Controlled Trials (CENTRAL), Ovid MEDLINE, Embase, PubMed, Latin American and Caribbean Health Sciences Literature Database (LILACS), Compendex, US National Institutes of Health Ongoing Trials Register, World Health Organization (WHO) International Clinical Trials Registry Platform was performed. Two independent reviewers screened titles and abstracts against the inclusion criteria, followed by the full-text articles. All included articles will be coded with the index test, input modality, population, target condition, validation, and reference standard. Finally, we will generate an interactive evidence map using EPPI Mapper.

Results:
A total of 18,895 records were identified from databases and registries. After removing duplicates, titles and abstracts for 14,171 records were screened. 13,034 records were excluded, 1137 records advanced to full-text screening, and 1118 studies were included in the review and 730 included in the current evidence map (Figure 1: Study flow diagram).
The majority of study reports described internal testing (n=661) of AI models using X imaging input data (648), with clinician graders as the reference standard (n,%), with glaucoma as the most frequent target condition (n=721) (Figure 2: Evidence Map).

Conclusions:
The majority of studies used the same participant data for AI model development and diagnostic performance evaluation (internal testing). Based on our preliminary results, most studies target glaucoma, use image data as input, and have a ground truth established by clinicians. Further research is required to analyze the overlap in datasets, assess for risk of bias, and fully understand AI's utility in glaucoma diagnosis.