INPLASY

INPLASY2025120099

doi: 10.37766/inplasy2025.12.0099

Received: 30 December 2025

Published: 30 December 2025

Diagnostic Accuracy of Deep Learning for Referable Age-Related Macular Degeneration: a Systematic Review and Meta-Analysis

Wang, TW; Luo, WT.

Corresponding author:
Ting Wei Wang

eltonwang1@gmail.com

Author Affiliation:
Taipei Veteran General Hospital.

ADMINISTRATIVE INFORMATION

Support - NR.

Review Stage at time of this submission - Completed but not published.

Conflicts of interest - None declared.

INPLASY registration number: INPLASY2025120099

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 30 December 2025 and was last updated on 30 December 2025.

INTRODUCTION

Review question / Objective To determine the diagnostic accuracy of deep learning algorithms applied to color fundus photographs for detecting referable age-related macular degeneration (rAMD) in adults. Primary objectives are to (1) pool sensitivity and specificity using a bivariate random-effects model, and (2) evaluate clinical utility using likelihood ratios. Secondary objectives are to compare DL performance with human graders using paired head-to-head evidence synthesized via a contrast-based network meta-analysis and to explore heterogeneity through prespecified subgroup analyses.

Condition being studied Age-related macular degeneration (AMD) is a progressive retinal disease and a major cause of central vision loss in older adults. This review focuses on “referable AMD,” generally corresponding to intermediate or advanced AMD that warrants specialist referral and/or closer monitoring due to higher risk of progression to late AMD and vision-threatening complications (e.g., neovascular AMD, geographic atrophy). Detection is performed using color fundus photography and graded using expert reference standards (commonly AREDS-based classification schemes). The condition is relevant to screening and referral workflows where accurate triage can reduce missed cases and avoid unnecessary specialist referrals.

METHODS

Participant or population Adults evaluated for referable age-related macular degeneration using color fundus photographs. Studies in screening, community, primary care, tertiary care, or mixed/public dataset settings will be eligible, provided adult participants are included and a reference standard based on expert human grading is available.

Intervention Index test: deep learning–based algorithms applied to color fundus photographs to classify referable AMD (binary or multi-class mapped to referable vs non-referable).

Comparator Reference standard: human expert grading (e.g., retinal specialists/reading centers) using accepted AMD grading schemes (commonly AREDS-based). For comparative analyses, human grader performance within the same cohort will serve as the comparator for head-to-head AI vs human evaluation.

Study designs to be included Diagnostic test accuracy studies (prospective or retrospective cohorts; external or internal validation) evaluating DL algorithms on color fundus photographs with sufficient data to derive 2×2 tables (TP, FP, TN, FN) for referable AMD.

Eligibility criteria Inclusion: (1) adult populations; (2) deep learning algorithm evaluated on color fundus photographs; (3) target condition includes referable AMD with a clear definition (e.g., intermediate/advanced AMD or equivalent mapping); (4) reference standard based on expert human grading; (5) diagnostic accuracy results enabling construction of 2×2 contingency tables.Exclusion: pediatric-only studies; non–deep learning methods; modalities other than color fundus photography (e.g., OCT, fluorescein angiography) as the primary index test; reviews/editorials; conference abstracts/proceedings; brief reports without extractable accuracy data; retracted studies; duplicate reports of the same dataset (most complete report retained).

Information sources Electronic databases: PubMed, Embase, Web of Science, IEEE Xplore. Additional sources: reference lists of included studies and relevant reviews. Where necessary, corresponding authors may be contacted for missing 2×2 data.

Main outcome(s) Primary outcomes: pooled sensitivity and specificity for DL detection of referable AMD from color fundus photographs, with 95% confidence intervals estimated using a bivariate random-effects model. Secondary main measures: positive and negative likelihood ratios (PLR, NLR) derived from pooled estimates.

Quality assessment / Risk of bias analysis Two reviewers will independently assess risk of bias and applicability using PROBAST with AI-adapted considerations (PROBAST-AI principles), covering participants, predictors, outcome (reference standard and blinding), and analysis (sample size, overfitting, leakage, and clustering). Each study will be rated low/unclear/high risk per domain and overall (highest-risk domain driving overall rating). Disagreements resolved by consensus.

Strategy of data synthesis We will compute study-level 2×2 tables and estimate pooled sensitivity and specificity using a bivariate random-effects logistic regression model. Forest plots and SROC curves will summarize performance. Likelihood ratios will be calculated and visualized using LR scattergrams and Fagan nomograms to illustrate post-test probability changes. For studies reporting paired AI and human grader results in the same cohort, we will synthesize relative sensitivity and relative specificity using a contrast-based approach to minimize bias from indirect comparisons; where appropriate, a network framework will be used to summarize paired comparisons across available architectures. Publication bias will be examined with Deeks' test. Analyses will be conducted in Stata (e.g., metadta/midas or equivalent commands).

Subgroup analysis Planned subgroup analyses include: architecture family (CNN vs transformer/other), economic status (World Bank classification), healthcare setting (primary vs tertiary), validation type (internal vs external), unit of analysis (patient vs eye), camera type (desktop vs smartphone vs ultrawide field), presence of multi-disease targets, reference standard criteria (AREDS vs other), adjudication approach, software certification/commercial status, vendor involvement, region, and handling of ungradable images.

Sensitivity analysis Sensitivity analyses will include restricting to studies with low overall risk of bias, excluding studies without external validation, excluding studies with unclear/atypical definitions of referable AMD, excluding studies with potential non-independence issues (e.g., both eyes per participant without adjustment) where identifiable, and assessing the influence of any single study via leave-one-out analyses when feasible.

Country(ies) involved Taiwan.

Keywords age-related macular degeneration; deep learning; fundus photography; diagnostic accuracy; screening; meta-analysis.

Contributions of each author
Author 1 - Ting Wei Wang.
Author 2 - Wei-Ting Luo.