INPLASY

Human Reader Expertise vs Deep Learning for Diabetic-Retinopathy Detection: A Systematic Review and Network Meta-analysis

INPLASY202580072
doi: 10.37766/inplasy2025.8.0072
Received: 23 August 2025
Published: 23 August 2025

Luo, WT; Tu, YK; Wu, YT; Wang, TW.

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: INPLASY202580072

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 23 August 2025 and was last updated on 23 August 2025.

INTRODUCTION

Review question / Objective In adults with diabetes undergoing fundus imaging, how does the diagnostic accuracy of deep learning (DL) algorithms compare with human graders (retina specialists, ophthalmologists, trained graders, or physicians) for detecting any, referable, or vision-threatening diabetic retinopathy?

Condition being studied Diabetic retinopathy (DR), a microvascular complication of diabetes and a leading cause of vision-threatening disease in working-age adults.

METHODS

Participant or population Adults (≥18 years) with type 1 or type 2 diabetes undergoing color fundus photography screening for DR.

Intervention Deep learning (DL) algorithms for automated DR detection.

Comparator Human graders: retina specialists, ophthalmologists, trained graders, or physicians.

Study designs to be included Diagnostic test accuracy studies, cross-sectional, cohort, randomized or non-randomized trials that allow construction of 2 × 2 diagnostic contingency tables.

Eligibility criteria Inclusion: Studies directly comparing DL with at least one human grader, using fundus imaging, reporting sufficient diagnostic accuracy metrics (sensitivity, specificity, AUC). Exclusion: Non-DL AI, pediatric populations, non-fundus imaging modalities, reviews/editorials, animal studies.

Information sources Electronic databases (PubMed, Embase, Cochrane Library, ClinicalTrials.gov, IEEE, arXiv, bioRxiv, medRxiv). Reference lists of included studies, and relevant grey literature.

Main outcome(s) Diagnostic accuracy: sensitivity, specificity, positive/negative predictive values, likelihood ratios, and AUC for DR detection at patient-level and eye-level.

Quality assessment / Risk of bias analysis Study quality assessed with QUADAS-AI across four domains (patient selection, index test, reference standard, flow/timing). Certainty of evidence graded using the GRADE framework (risk of bias, indirectness, inconsistency, imprecision, publication bias).

Strategy of data synthesis 2 × 2 contingency tables reconstructed per study. Hierarchical bivariate random-effects models used for pooled sensitivity/specificity. Network meta-analysis (contrast-based) for comparing DL with multiple human grader categories. Heterogeneity explored with meta-regression.

Subgroup analysis By reader expertise (retina specialist, ophthalmologist, trained grader, physician), economic setting (high vs middle income), DR type (any, referable, vision-threatening), pupil dilation status, imaging modality, and handling of ungradable images.

Sensitivity analysis Excluding studies with high risk of bias, studies without external validation, vendor-involved studies, and conference abstracts.

Country(ies) involved Taiwan.

Keywords Artificial intelligence; deep learning; diabetic retinopathy; screening; systematic review; meta-analysis; diagnostic accuracy.

Contributions of each author

Author 1 - Wei-Ting Luo.
Email: luomaggie1@gmail.com
Author 2 - Yu-Kang Tu.
Email: yukangtu@ntu.edu.tw
Author 3 - Yu-Te Wu.
Email: ytwu@nycu.edu.tw
Author 4 - Ting Wei Wang.
Email: eltonwang1@gmail.com