International Platform of Registered Systematic Review and Meta-analysis Protocols

INPLASY

INPLASY202620069

doi: 10.37766/inplasy2026.2.0069

Received: 23 February 2026

Published: 23 February 2026

Comparative Accuracy of AI Systems for Referable Diabetic Retinopathy Screening: A Systematic Review and Bayesian Network Meta-analysis

Wang, TW; Fang, WC; Luo, WT; Tu, YK; Wu, YT; Chou, YB.

Corresponding author:
Ting Wei Wang

eltonwang1@gmail.com

Author Affiliation:
Taipei Veterans 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: INPLASY202620069

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 23 February 2026 and was last updated on 23 February 2026.

INTRODUCTION

Review question / Objective In adults undergoing diabetic retinopathy screening using color fundus photographs, to compare commercially available, regulator-approved deep-learning automated retinal image analysis systems (ARIAS) with each other for detecting referable diabetic retinopathy (rDR) and screening-relevant consequences (missed rDR and referral burden). We will synthesize head-to-head or multi-algorithm diagnostic accuracy evidence using Bayesian hierarchical bivariate network meta-analysis, estimating pooled sensitivity/specificity and probabilistic rankings (posterior rank probabilities and SUCRA).

Condition being studied Diabetic retinopathy is a microvascular complication of diabetes and a major cause of vision impairment. Vision-threatening disease can be asymptomatic until advanced stages; therefore, screening programs use retinal fundus photographs graded using standardized DR severity frameworks to identify referable DR (rDR), a threshold indicating need for ophthalmic evaluation and possible treatment. rDR commonly corresponds to moderate non-proliferative DR or worse and/or diabetic macular edema depending on the program definition. Screening pathways vary by setting (primary care, secondary/tertiary clinics, national programs), camera type, pupil dilation, field strategy, and handling of ungradable images. These factors influence accuracy and referral burden and are critical for evaluating deployable AI systems.

METHODS

Participant or population Adults (≥18 years) undergoing diabetic retinopathy screening or evaluation using color fundus photographs in any setting (primary care screening, secondary/tertiary clinics, national screening programs). Mixed-age

cohorts will be eligible if ≥85% are adults or if adult-only results allow construction of 2×2 tables.

Intervention Commercially available, regulator-approved deep learning automated retinal image analysis systems (ARIAS) applied to color fundus photographs for detection of referable diabetic retinopathy (rDR).

Comparator Other commercially available, regulator-approved deep learning ARIAS evaluated head-to-head in the same study population. (Where reported, human grading will be the reference standard rather than a comparator for the network.).

Study designs to be included Prospective or retrospective diagnostic accuracy studies with head-to-head or multi-algorithm evaluation of ≥2 eligible commercial deep learning ARIAS in the same population, reporting or permitting reconstruction of TP/FP/FN/TN. Peer-reviewed full-text articles only for quantitative synthesis.

Eligibility criteria Inclusion: (1) Adults undergoing DR screening/evaluation using color fundus photographs; (2) diagnostic accuracy studies evaluating ≥2 commercially available, regulator-approved deep learning ARIAS head-to-head in the same population; (3) human grading reference standard using recognized DR classification schemes; (4) sufficient data to reconstruct TP/FP/FN/TN for rDR; (5) prospective or retrospective design; (6) peer-reviewed full text available by 29 Jan 2026.Exclusion: non-head-to-head single-algorithm studies; non-deep learning approaches; non-fundus modalities (OCT/OCTA); non-image risk models; algorithm-only reference standards; development-only reports without distinct external/hold-out validation; preprints/abstracts without peer-reviewed full text by cutoff. Non-overlapping strata with separable 2×2 tables will be extracted as independent analytic units. Patient-level and eye-level datasets will be analyzed separately.

Information sources PubMed, Embase, and ClinicalTrials.gov from inception to 29 January 2026, plus citation chasing of included studies and relevant reviews. Authors may be contacted if required data for 2×2 reconstruction are missing.

Main outcome(s) Primary outcomes are diagnostic accuracy for detecting referable diabetic retinopathy (rDR): sensitivity and specificity for each ARIAS, based on reconstructed 2×2 tables (TP/FP/FN/TN). Comparative outcomes include relative ranking of ARIAS using posterior rank probabilities and SUCRA (based on Youden index) within the connected evidence network. Patient-wise and eye-wise outcomes will be synthesized separately.

Quality assessment / Risk of bias analysis Risk of bias and applicability will be assessed using QUADAS-2 across four domains (patient selection, index test, reference standard, flow/timing). AI-specific concerns will be informed by PROBAST-AI (e.g., leakage, spectrum effects, selective reporting, post hoc threshold tuning, and handling of ungradable images). Two reviewers will assess independently; disagreements will be resolved by discussion with a third reviewer.

Strategy of data synthesis We will reconstruct TP/FP/FN/TN and synthesize diagnostic accuracy using a Bayesian hierarchical bivariate network meta-analysis with binomial likelihoods and logit links, jointly modeling sensitivity and specificity and their correlation via correlated study-level random effects. Additional study-by-test random effects will be included to accommodate residual heterogeneity among algorithms within studies. Weakly informative priors will be used for variance components and correlation. Posterior inference will be obtained by MCMC in JAGS via R (R2jags). Results will be reported as posterior means and 95% credible intervals. Rankings will use posterior rank probabilities and SUCRA based on Youden index, with marginal rankings for sensitivity and specificity. Patient-wise and eye-wise networks will be analyzed separately. Convergence will be assessed using trace/density plots and Gelman–Rubin diagnostics.

Subgroup analysis Prespecified subgroup/meta-regression covariates may include setting (screening vs clinic), camera type (desktop/handheld/smartphone), mydriasis strategy (yes/no/mixed), field strategy (2/3/≥4 fields), reference standard pathway (photograph grading vs clinically anchored examination), and handling of ungradable images (excluded vs classified as positive/other). Analyses will be conducted when data support sufficient strata.

Sensitivity analysis Sensitivity analyses may include restricting to studies at low risk of bias in key QUADAS-2 domains; excluding studies with unclear/high concern regarding flow/timing or ungradable handling; restricting to peer-reviewed prospective studies; and evaluating influence of large studies on network estimates. Alternative ranking metrics (e.g., sensitivity-only or specificity-only) will be reported to contextualize SUCRA results under high heterogeneity.

Country(ies) involved Taiwan.

Keywords diabetic retinopathy; screening; artificial intelligence; deep learning; automated retinal image analysis; network meta-analysis; Bayesian; diagnostic accuracy.

Contributions of each author
Author 1 - Ting-Wei Wang.
Author 2 - Wei-Chieh Fang.
Author 3 - Wei-Ting Luo.
Author 4 - Yu-Kang Tu.
Author 5 - Yu-Te Wu.
Author 6 - Yu-Bai Chou.