INPLASY

INPLASY202630009

doi: 10.37766/inplasy2026.3.0009

Received: 2 March 2026

Published: 3 March 2026

Corresponding author:
Ting Wei Wang

eltonwang1@gmail.com

Author Affiliation:
Taipei Veteran General Hospital.

Deep learning, human readers, and human–AI collaboration for diabetic retinopathy screening: a Bayesian network meta-analysis of diagnostic test accuracy

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

ADMINISTRATIVE INFORMATION

Support - NR.

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

Conflicts of interest - None declared.

INPLASY registration number: INPLASY202630009

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 3 March 2026 and was last updated on 3 March 2026.

INTRODUCTION

Review question / Objective Primary objective: To simultaneously compare and probabilistically rank the diagnostic accuracy (sensitivity and specificity) of three deployment strategies—AI-alone, human-only, and human–AI collaboration—for diabetic retinopathy screening using Bayesian network meta-analysis of diagnostic test accuracy (DTA-NMA).
Secondary objectives: (1) To explore sources of heterogeneity through pre-specified hierarchical meta-regression on 26 theory-driven covariates; (2) To generate context-specific, equitable deployment recommendations for national screening programmes, with particular focus on low- and middle-income settings.

Condition being studied Diabetic retinopathy (DR) – a microvascular complication of diabetes mellitus that progresses from mild non-proliferative changes to vision-threatening proliferative disease and macular oedema. Early detection through regular fundus screening followed by timely laser, anti-VEGF, or surgical intervention can prevent severe vision loss.Diabetic retinopathy affects an estimated 103 million people globally and is projected to reach 160 million by 2045, remaining the leading cause of preventable blindness among working-age adults. When linked to timely treatment, regular screening can substantially reduce the risk of severe vision loss, yet real-world adherence is starkly unequal (66 % in high-income countries vs 39 % in LMICs). Workforce shortages, geographic inaccessibility, and high costs exacerbate this inequity, resulting in low coverage and delayed diagnosis precisely where consequences are most severe. Screening programmes worldwide therefore face a critical systems-level decision: they must select a deployment workflow that realistically matches local capacity, budget constraints, and follow-up infrastructure.
More than 15 pairwise meta-analyses have demonstrated high diagnostic accuracy of deep-learning systems, yet these studies assess only one strategy in isolation and cannot rank the three available options (AI-alone, human-only, human–AI

collaboration). No previous synthesis has simultaneously compared and ranked these strategies or provided context-specific guidance. A Bayesian DTA-NMA is uniquely suited to close this gap because it enables head-to-head comparisons through a connected evidence network, propagates uncertainty transparently, produces probabilistic rankings with credible intervals, and accommodates imperfect reference standards, grader variability, and spectrum bias. By incorporating 26 pre-specified covariates through hierarchical meta-regression with shrinkage priors, this protocol will deliver practical, transportable recommendations to support equitable and sustainable blindness prevention worldwide.

METHODS

Participant or population Adults (≥18 years) with diabetes mellitus (type 1 or 2) undergoing diabetic retinopathy screening with colour fundus photography in any clinical or community setting. Studies of children, pregnant women with gestational diabetes only, or non-fundus imaging modalities will be excluded.

Intervention Three deployment strategies for DR screening:• AI-alone (fully automated grading with direct referral/recall rules)• Human-only (manual grading by trained readers or ophthalmologists without AI)• Human–AI collaboration (integrated workflows; mapped to three pre-specified archetypes: AI-triage + human over-read, human-first + AI second-reader/arbitrator, or structured disagreement resolution).

Comparator The other two deployment strategies (head-to-head comparison via network meta-analysis).

Study designs to be included Prospective and retrospective diagnostic accuracy studies, pivotal validation trials, and real-world implementation studies that report sufficient data (2×2 contingency tables or sensitivity/specificity with 95 % CI) to reconstruct diagnostic performance for at least one of the three strategies against a reference standard (usually ophthalmologist grading or ETDRS photography).

Eligibility criteria Inclusion: Studies evaluating any of the three deployment strategies with extractable accuracy data.Exclusion: Studies evaluating only algorithm development without external validation; studies without a human or AI comparator; animal or in-vitro studies; conference abstracts without full data; duplicate publications (most recent retained).

Information sources Electronic databases (PubMed, Embase, Web of Science), and manual reference checking. Authors of eligible studies will be contacted for missing data.

Main outcome(s) Pooled sensitivity and specificity (with 95 % credible intervals) for referable DR and vision-threatening DR at patient and eye level for each deployment strategy; probabilistic ranking of strategies (surface under the cumulative ranking curve and rank probabilities).

Quality assessment / Risk of bias analysis QUADAS-2 tool (with AI-specific extensions for reference standard and flow/timing domains). Risk of bias and applicability concerns will be summarised graphically and narratively.

Strategy of data synthesis Bayesian bivariate network meta-analysis using the metamiss or Bayesian bivariate model framework (R/Stan or WinBUGS). We will model sensitivity and specificity jointly, account for within-study correlation, and produce probabilistic rankings. Heterogeneity will be explored through hierarchical meta-regression on 26 pre-specified covariates with shrinkage priors. Inconsistency will be assessed by node-splitting; publication bias via comparison-adjusted funnel plots. All analyses will follow PRISMA-DTA and PRISMA-NMA extensions.

Subgroup analysis Heterogeneity will be explored through hierarchical meta-regression on 26 pre-specified covariates with shrinkage priors.

Sensitivity analysis Exclusion of high risk-of-bias studies, studies with industry funding, studies using non-adjudicated reference standards, and studies with >20 % ungradable images.

Country(ies) involved Taiwan.

Keywords diabetic retinopathy; screening; artificial intelligence; deep learning; network meta-analysis; diagnostic test accuracy; Bayesian; deployment strategy; human-AI collaboration.

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-Bai Chou.
Author 6 - Yu-Te Wu.