Online Resource 2: TRIPOD+AI reporting checklist and CLAIM 2024 / STARD-AI topic map

Retinal foundation-model versus ImageNet pretraining for deep learning grading of diabetic retinopathy: an external validation study

Pravinraj Ethiraj — Department of Artificial Intelligence and Data Science, Chennai Institute of Technology, Chennai, Tamil Nadu, India — pravinraje.aids2024@citchennai.net

International Ophthalmology

Guideline: Collins GS et al. TRIPOD+AI statement. BMJ 2024;385:e078378. The study is a development study with external validation of existing model architectures and no model updating after external testing. Topics below are abridged; item wording follows the official TRIPOD+AI checklist. Locations refer to the main article (section headings) and Online Resources (OR).

Topic (abridged)	Reported	Location
Title and abstract		
Title identifies development/validation, population and outcome	Yes	Title
Structured abstract (objectives, data, methods, results, conclusions)	Yes	Abstract (Purpose, Methods, Results, Conclusion)
Introduction		
Background, clinical context and rationale	Yes	Introduction, paragraphs 1–3
Objectives and hypothesis	Yes	Introduction, paragraph 4
Methods — data		
Data sources for development and evaluation; retrospective design	Yes	Methods: Study design; Datasets and reference standards; Table 1
Dates of data access	Yes	References (APTOS access date); OR1 S1
Setting (countries, sites, devices)	Yes	Table 1
Eligibility (gradability) and image counts	Yes	Datasets (Messidor-2 1,744 of 1,748 gradable; IDRiD 516); Table 1
Previously published datasets	Yes	Datasets
Methods — outcome and reference standard		
Outcome definition (five-level ICDR grade; referable DR = grade ≥ 2)	Yes	Datasets and reference standards
Reference standard per dataset; adjudication; label harmonization	Yes	Table 1; Datasets and reference standards
Blinding of reference-standard graders	Not applicable (pre-graded public data)	Strengths and limitations
Methods — input data and preprocessing		
Image preprocessing, identical for all datasets	Yes	Preprocessing
Sample size (fixed by public releases)	Yes	Table 1; Strengths and limitations
Missing data / label quality (audit of a third-party IDRiD copy)	Yes	Datasets; OR1 S1
Methods — model development		
Architectures and pretraining	Yes	Models and training; OR1 Table S1
Training procedure (loss, optimizer, schedule, augmentation, early stopping)	Yes	Models and training; OR1 Table S1
Differences between model recipes; label-smoothing and crossed-recipe controls	Yes	Models and training; OR1 Table S1; Results (Exploratory: matched-recipe and label-smoothing controls); OR1 S7
Repeated training runs (seeds)	Yes	Models and training; OR1 Table S1; Table 2; OR1 S6
Hyperparameter and checkpoint selection on internal validation only	Yes	External testing and leakage controls; OR1 S3
How predictions are generated (softmax; argmax; probabilities retained)	Yes	Outcomes and statistical analysis; Uncertainty and selective referral
Methods — evaluation		
Performance measures and rationale (QWK primary)	Yes	Outcomes and statistical analysis
Primary estimand, confirmatory family, multiplicity	Yes	Outcomes and statistical analysis
Changes to the analysis plan, with original results	Yes	Methods: Changes to the analysis plan; Table 2; OR1 S0
Clustering of images (Messidor-2 eye pairs)	Yes (examination-level bootstrap with the official ADCIS examination file)	Outcomes and statistical analysis; Strengths and limitations; OR1 S4
Calibration	Yes	Outcomes and statistical analysis; Results (Exploratory: calibration); OR1 S5
Uncertainty and selective referral, compared with simple confidence	Yes	Uncertainty and selective referral; Results; Figure 4; OR1 S9
Explainability evaluated quantitatively against expert lesion masks, with controls	Yes	Saliency localization against expert lesion masks; Results; Figure 5; OR1 S10
Case mix and no-DR/mild boundary analysis (descriptive)	Yes	Case mix and the no-DR/mild boundary; Figure 3; OR1 S8
Fairness / subgroup analysis by demographic factors	Not possible (no demographic data in the public releases)	Strengths and limitations

Topic (abridged)	Reported	Location
Open science		
Funding; competing interests	Yes	Statements and Declarations (Funding; Competing interests)
Protocol / registration	Not registered (retrospective analysis of public data)	—
Data availability	Yes	Statements and Declarations (Data availability); OR3
Code availability	Yes	Statements and Declarations (Code availability); OR4
Results		
Image flow and characteristics per dataset	Yes	Table 1; Figure 1
Model specification	Yes	OR1 Table S1; OR1 S3 (checkpoint hashes); OR4
Internal and external performance with 95% CI, including sensitivity, specificity, PPV and NPV	Yes	Table 3; OR1 S5
Between-model difference in drop with CI and adjusted p	Yes	Table 2; Figure 2; OR1 S6
Referable DR, case mix, calibration, uncertainty, localization	Yes	Results; Table 3; Figures 3–5; OR1 S8–S10
Sensitivity analyses, flagged as such	Yes	Results (Matched-recipe controls; Additional analyses); OR1 S6, S7, S11
Discussion		
Interpretation and comparison with prior work	Yes	Discussion
Strengths and limitations	Yes	Strengths and limitations
Clinical implications without overstatement	Yes	Discussion (Strengths and limitations; Conclusions)
Other		
Ethics and consent	Yes	Methods: Ethics statement; Statements and Declarations (Ethics approval; Consent)
Use of generative AI	Yes	Methods: Software, reproducibility and use of generative AI

CLAIM 2024 and STARD-AI: where the main topics are reported

CLAIM 2024 (Tejani AS et al., Radiol Artif Intell 2024;6:e240300) and STARD-AI (Sounderajah V et al., Nat Med 2025;31:3283–3289) cover overlapping topics. The locations below map those topics to the article; the official checklists should be consulted for exact item wording.

Topic	Location
Study design, intended use and clinical role (retrospective external evaluation; not a clinical deployment study)	Introduction; Methods: Study design; Limitations
Data sources, inclusion/exclusion, image counts and flow	Datasets; Table 1; Figure 1
Reference standard: definition, graders, adjudication, rationale	Table 1; Datasets and reference standards
Data partitions and independence of test sets	Datasets; External testing and leakage controls; OR1 S2, S4
Preprocessing and de-identification	Preprocessing; Ethics statement
Model description, initialization and training details	Models and training; OR1 Table S1; OR4
Metrics, statistical tests, confidence intervals, multiplicity	Outcomes and statistical analysis
External testing; failure analysis; subgroup by disease grade	Results; Figure 3; OR1 S8
Calibration and uncertainty	Results; Figure 4; OR1 S5, S9
Explainability evaluation	Results; Figure 5; OR1 S10
Limitations, generalisability, potential bias	Discussion; Strengths and limitations
Data, code and model availability	Statements and Declarations (Data availability); OR1 S3; OR3; OR4