Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Automated triage of open globe injury from external-eye photographs for low-resource settings

Benton Chuter

benton.chuter@hitchcock.org

Dartmouth–Hitchcock Medical Center

Fae B. Kayarian

Dartmouth–Hitchcock Medical Center

Min Young Kim

University of Tennessee Health Science Center

Bailey R. Millis

Dartmouth–Hitchcock Medical Center

Fabliha Mukit

University of Tennessee Health Science Center

Timothy Lee

Walter Reed National Military Medical Center

Paul Chong

Walter Reed National Military Medical Center

Grant Justin

Walter Reed National Military Medical Center

Taylor Miller

Dartmouth–Hitchcock Medical Center

Jon Marshall

Dartmouth–Hitchcock Medical Center

Monica M. Jablonski

University of Tennessee Health Science Center

William G. Gensheimer

Dartmouth–Hitchcock Medical Center

Article

Keywords: open globe injury, ocular trauma, triage, foundation models, external-eye photography, low-resource settings, combat ocular trauma, digital health

Posted Date: August 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10386119/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Open globe injury is a vision-threatening emergency whose outcome depends on rapid escalation to surgical repair, yet in rural, military, and other low-resourced settings the escalate-or-not decision is usually made by non-ophthalmologists without access to a slit lamp. We present a smartphone-deployable triage tool that detects likely open globe injury from an ordinary external-eye photograph and runs fully offline, with no image or protected health information leaving the phone. From a heterogeneous, multi-source, real-world cohort (1,305 images, 58 open globe positives) with four-reviewer adjudicated labels, we trained and evaluated vision backbones under a de-duplicated, leakage-free split with a frozen held-out test (261 images, 12 positives). A fine-tuned ViT-B reached AUROC 0.934 on the frozen held-out test (0.930 mean across ten seeds; Supplementary Fig. S3); a frozen-feature probe and ensemble, with no backbone fine-tuning, nearly matched it (VisionFM probe 0.906, ensemble 0.908) while running on edge hardware, with a conservative cross-validation estimate of 0.88 to 0.89. We package the model into a working iOS and Android application and map its use across military, community, emergency department, and telemedicine pathways. To our knowledge, this is the first automated open globe injury triage signal from external-eye photographs, delivered as a deployable, offline tool for low-resourced settings.

Introduction

Ocular trauma is a leading cause of monocular blindness worldwide, with open globe injury (OGI), defined as a full-thickness wound of the cornea and/or sclera, representing one of the most time-sensitive forms of ocular trauma. In OGI, primary surgical repair should be performed within 24 hours of injury to reduce risk of endophthalmitis and permanent vision loss.[1,2] The initial assessment of ocular trauma is frequently performed by non-ophthalmologists who must determine whether urgent ophthalmologic consultation is indicated.[3] In active military conflict zones, initial assessment and the evacuation decisions are often determined by medics, corpsmen, physician assistants, or non-ophthalmologist clinicians[4–12]. The Joint Trauma System (JTS) Clinical Practice Guidelines (CPG) recommend evacuation of “all vision-threatening injuries so that they receive treatment by an eye surgeon within 24 hours, if possible.”[13]

The global burden of ocular trauma disproportionately affects patients in low- and middle-income countries (LMICs). In 1998, the World Health Organization (WHO)-affiliated estimates placed eye injury at roughly 55 million activity-limiting injuries annually, including approximately 200,000 OGIs requiring hospitalization, with about 1.6 million people blind from injury and a further 19 million with unilateral vision loss.[14] Recent Global Burden of Disease estimates are consistent with earlier reports, indicating approximately 59.9 million incident eye injuries worldwide in 2019 and demonstrating persistent disparities across the socioeconomic development gradient.[15,16] Hospital series from LMICs find OGI to be a large share of presenting ocular trauma, roughly 30% in Malawi to roughly 56% in Ghana and Malaysia, affecting predominantly young, working-age males and children, with poor visual outcomes[17–

[20]. Delay in ophthalmic care for OGI is common and of great clinical consequence. In one pediatric cohort, only 39% of children with OGI presented to a clinical setting within 24 hours, and earlier presentation tripled the odds of a favorable visual outcome.[19] One structural driver in the existing literature appears to be a global ophthalmologist shortage: ophthalmologist density is 76 per million in high-income versus 3.7 per million in low-income countries, a greater than 20-fold difference. Approximately 17% of the world's population, across 132 countries, has access to under 5% of the world's total practicing ophthalmologists [21].

Automated decision support that runs on an external-eye photograph rather than a slit lamp could enable a frontline provider to make the “escalate-or-not call” more accurately and efficiently without an ophthalmic specialist present. While deep learning has been validated for fundus-based screening in disease states including diabetic retinopathy and glaucoma, our team is aware of no prior automated approach to OGI detection from external-eye photographs. Here, we establish that this signal exists and is learnable from the uncontrolled, real-world imagery that such settings produce (Fig. 1), and we benchmark a range of general-purpose and domain-specific vision backbones to characterize it.**Figure 1.** Representative external-eye photographs by category. License-clear external-eye photographs (open globe, indeterminate, non-open globe) shown for visual review of

input-image quality and the range of real-world imagery. Figure 1 image credits (all from

Wikimedia Commons, cropped to the ocular region): open globe, "Penetrating eye injury" by Bobjgalindo (CC BY-SA 4.0), "Globe rupture" by S. Bhimji MD (CC BY 4.0), "Blunt ocular trauma - iris protrusion" by Maria Sieglinda von Nudeldorf (CC BY-SA 4.0); indeterminate, "Hyposphagma rechts Blick leicht nach unten" by Andrea Kamphuis (CC BY-SA 4.0), hyphema image from Okunuki et al., BMC Ophthalmol. 2014;14:74 (CC BY 4.0), "Hyphema - occupying half of anterior chamber of eye" by Rakesh Ahuja MD (CC BY-SA 2.5); not open globe, “Black eye as result of contusion" by Mariano-J (CC BY-SA 3.0), "Human cornea with abrasion highlighted by fluorescein staining" by James Heilman MD (CC BY-SA 3.0), “Subconjunctival hemorrhage, small and temporal" by Iceclanl (CC BY-SA 4.0). This figure contains third-party images licensed under their own Creative Commons terms as credited above; they are not covered by the article's CC BY license.

Results

Cohort and task

A cohort was built from a multi-tier web scrape of 11,146 external-eye images; automated filtering and a first-pass screen for image content, licensing, and gradability kept 1,657. A perceptual-hash and pixel-correlation audit then removed 52 duplicate images, and a further 300 images were removed during domain-expert grading, yielding a final cohort of 1,305 unique images (58 OGI positives) with four-reviewer adjudicated labels. The task is binary detection of "OGI" versus "everything else" (closed globe trauma, suspected-but-not-open, and non-trauma). Images pooled across source tiers were split by

perceptual-hash cluster-grouped, label-stratified holdout (Fig. 2) into a train+validation (train + val) set (1,044 images, 46 positives) for all model fitting and a frozen held-out test (261 images, 12 positives) touched only for final scoring. The reported split is de-duplicated and leakage-free. All reported numbers are on the frozen test set unless otherwise stated, with 2,000-resample bootstrap 95% confidence intervals (Table 1).**Figure 2.** Study design. A single leakage-safe cohort (dedup-cluster-grouped, label-stratified) is evaluated across five backbones, two representation modes (frozen-feature linear probe and fine-tuning), two label tasks (binary open globe-vs-rest and tertiary), and six combination strategies, scored by internal grouped cross-validation and a frozen held-out test with bootstrap and DeLong inference.

Table 1 Cohort composition and split.

Table 1
Cohort composition and split, with open-globe positives by split. Train + val comprises 1,044 images (46 open-globe positives); the frozen held-out test comprises 261 images (12 positives); the total curated cohort is 1,305 images (58 positives).

Split	Images	Open-globe positives
Train + val (model fitting)	1,044	46
Frozen held-out test	261	12
Total curated cohort	1,305	58

A compute-light frozen probe nearly matches the best fine-tuned model

A fine-tuned ImageNet ViT-B/16 was the strongest model on the frozen test (AUROC 0.934; stable at 0.930 across ten seeds). The ophthalmology foundation model VisionFM, used only as a frozen feature extractor, was the best single probe (0.906, 95% CI 0.843 to 0.957). Other backbones trailed (ImageNet ViT-B/16 probe 0.779; DINOv2 ViT-B/14 probe 0.701; ConvNeXt-Tiny probe 0.794; EfficientNet-B0 probe 0.888, fine-tune 0.742). For deployment, a frozen probe on domain features captures most of the signal with no backbone fine-tuning, well suited to compute-light edge inference (Table 2).

Table 2
Per-backbone performance on the frozen held-out test.

Backbone	Linear-probe AUROC [95% CI]	Probe sens@90spec	Fine-tune AUROC	Fine-tune sens@90spec
VisionFM (ophthalmology FM)	0.906 [0.843, 0.957]	0.667	0.880	0.583
ImageNet ViT-B/16	0.779 [0.564, 0.950]	0.667	0.934	0.917
DINOv2 ViT-B/14	0.701 [0.508, 0.868]	0.417	NA	NA
ConvNeXt-Tiny	0.794 [0.627, 0.926]	0.500	NA	NA
EfficientNet-B0	0.888 [0.785, 0.960]	0.583	0.742	0.083

Table 2. Per-backbone performance on the frozen held-out test (frozen linear probe versus end-to-end fine-tune). AUROC with 2,000-resample bootstrap 95% confidence intervals; sensitivity at > = 90% specificity. NA denotes fine-tune not run for that backbone.

Model combination does not exceed the single backbone

Our team compared six combination strategies, fitting every weight and meta-learner on train + val out-of-fold predictions only. Early feature fusion (concatenated frozen embeddings followed by class-balanced logistic regression) gave the best point estimate (AUROC 0.912, 95% CI 0.803 to 0.979; sensitivity 0.75 at > = 90% specificity; expected calibration error 0.036); a greedy meta-ensemble collapsed onto feature fusion (Fig. 3). The hybrid ensemble gave the best calibration (ECE 0.025) (Fig. 4). No combination significantly beat the single VisionFM probe (DeLong [22] fusion vs VisionFM: z = 0.14, p = 0.89), as expected with only 12 test positives (Table 3).**Figure 3.** Discrimination and calibration on the frozen held-out test. (A) Receiver-operating-characteristic curves for the leading models (feature fusion, VisionFM probe, soft-vote ensemble, ImageNet ViT-B/16). (B) Precision-recall curves (positive prevalence 0.046 marked as baseline). (C) Calibration (reliability) curves for the three best-calibrated models.

Table 3 Model-combination strategies on the frozen held-out test.

Table 3
Model-combination strategies on the frozen held-out test. AUROC (95% CI), PR-AUC, sensitivity at > = 90% and > = 95% specificity, and expected calibration error (ECE). All weights and meta-learners were fit on train + val out-of-fold predictions only.

Method	AUROC [95% CI]	PR-AUC	sens@90spec	sens@95spec	ECE
Feature fusion (concat embeddings)	0.912 [0.803, 0.979]	0.487	0.75	0.583	0.036
Hybrid (greedy over bases+fusion+stacking)	0.909 [0.830, 0.968]	0.475	0.583	0.500	0.025
Soft-vote ensemble	0.908 [0.827, 0.972]	0.489	0.583	0.583	0.033
Rank-mean ensemble	0.890 [0.760, 0.974]	0.476	0.667	0.667	0.454
Stacking (logistic meta-learner)	0.904 [0.822, 0.969]	0.450	0.583	0.583	0.281
Greedy (Caruana) ensemble	0.904 [0.837, 0.958]	0.410	0.583	0.417	0.035

Internal versus held-out estimates

Held-out AUROCs exceeded the train + val cross-validation estimates (for example, VisionFM 0.869 CV to 0.906 test), consistent with a small, favorable test draw. The cross-validation values (0.88 to 0.89) is a conservative alternate primary estimate, with the frozen-test number as held-out confirmation, given the wide test intervals (Table 4).

Table 4
Internal cross-validation versus held-out AUROC.

Model	Train + val CV AUROC [CI]	Frozen-test AUROC [CI]
Greedy ensemble	0.892 [0.840, 0.936]	0.904 [0.837, 0.958]
VisionFM	0.869 [0.820, 0.915]	0.906 [0.843, 0.957]
DINOv2	0.830 [0.764, 0.891]	0.701 [0.508, 0.868]
ImageNet ViT-B	0.780 [0.697, 0.855]	0.779 [0.564, 0.950]

Table 4. Internal grouped cross-validation versus frozen held-out AUROC (95% CI).

Robustness and shortcut checks

Because the cohort is web-scraped, we audited for source-driven shortcuts (Supplementary Table S1). Leave-one-tier-out and tier-stratified cross-validation AUROC stay above chance across all three source tiers, so the signal is not confined to a single source. Excluding face-containing images leaves AUROC unchanged, and source tier is not strongly decodable from the embeddings (cross-validation accuracy 0.925 at the 0.936 majority baseline). The image model uses no captions at inference.

Discussion

Open globe injury is detectable by a smartphone-based AI tool from an external-eye photograph at performance in the range of well-regarded fundus-based screening models built on larger datasets, despite a small positive count and deliberately uncontrolled imagery. To our knowledge, this is the first automated approach to this task with this modality, and it defines a benchmark the field can build on for clinically meaningful benefits.

Our deployment target includes patients and clinicians in low-resource, rural, and military settings. A photograph-based triage signal that runs on smartphones (iOS, Android) lets a non-ophthalmologist decide “escalate-and-shield-now” versus routine referral to outpatient ophthalmology for non-open globe ocular trauma, which could help reduce inequities in emergency eye care. Our design choices are oriented toward implementation rather than benchmark performance alone. This includes deployment-representative training data, a frozen leakage-free test, calibration suited to a referral threshold, and a compute-light frozen-feature probe deployable on edge hardware. We report decision-curve net benefit across a range of referral thresholds (Supplementary Fig. S2). We illustrate this deployment path across representative eye-care settings (Fig. 5) and in a working, fully offline smartphone application (Fig. 6). Because inference runs entirely on the smartphone with no connectivity, no cloud service, and no image or protected health information leaving the device, the tool fits settings with little digital infrastructure and sidesteps the data-governance barriers that stall many digital-health deployments. A single, mid-range smartphone is the only hardware required, and the frozen-feature probe not requiring backbone fine-tuning keeps the compute footprint small enough for older devices. In a non-specialist workflow, the output is a single escalate-and-shield versus routine-referral call rather than a diagnosis, which is the clinical decision a frontline provider needs to make. The heterogeneous, multi-source web-scraped cohort is a strength of this dataset, as it samples the variability (e.g. device, lighting, angle, background) of in-field imaging far better than a curated single-device dataset would.

Several limitations of our model remain. First, with only 12 open globe positives in the held-out test, confidence intervals are wide (Table 2), and the top methods cannot be separated. Second, although the binary open-globe target inter-grader agreement was high (raw agreement 97%, Gwet AC1 0.96) and all disagreements were resolved by expert adjudication, additional independent reads would harden the ground truth. Third, open globe-only is a first-pass triage signal. This tool lacks the ability to complete a full ocular-trauma assessment, since a large share of vision-threatening trauma is not open globe

(oculoplastic, orbital fracture, traumatic optic neuropathy, hyphema, retinal injury). Fourth, prospective field deployment and referral-workflow integration remain future work, alongside external validation on a domain-matched low resource-setting clinical cohort. While this working, offline smartphone application (Fig. 6) demonstrates implementation, its field evaluation is currently prospective.

Overall, a domain ophthalmology foundation model, used even as a frozen feature extractor, detects open globe injury from external-eye photographs well enough to reliably perform as an automated non-specialist triage tool. The path from this benchmark to a deployable, equity-oriented digital triage aid is concrete and has immense clinical potential.

Methods

Data collection and labeling

External-eye images were assembled by a multi-tier web scrape (open-access, restricted, teaching, and social sources) plus a curated scrape tier, and de-duplicated by perceptual hashing into clusters. Two reviewers screened each image to keep or reject it for eye content and gradability (J.M., T.M.) (Supplementary Fig. S1). Four graders labeled the kept images for open globe injury (F.M., T.L., P.C., G.J.), and an expert adjudicated all disagreements (W.G.G.), most of which traced to a single grader. The multi-class reads were collapsed to a binary target (open globe vs rest), and labels are the adjudicated consensus.

Splits

We used a perceptual-hash cluster-grouped, label-stratified holdout (seed 20260621): a train+valid set (1,044 images, 46 positives) for all model fitting and a frozen held-out test (261 images, 12 positives) for final scoring only. No near-duplicate cluster straddles the split, which was verified to be leakage-free.

Models and training

Backbones spanned a domain ophthalmology foundation model (VisionFM [23]) and general-purpose encoders (ImageNet ViT-B/16 [24], DINOv2 ViT-B/14 [25], ConvNeXt-Tiny [26], EfficientNet-B0 [27]). Two paradigms were evaluated: (1) a frozen-feature linear probe, a class-balanced logistic-regression classifier trained on the labeled training embeddings with the backbone held fixed, and (2) end-to-end fine-tuning, which also updates the backbone. Eval-time preprocessing was Resize(256), CenterCrop(224), then ImageNet normalization. VisionFM weights are released under CC-BY-NC 4.0 (academic use). Models were implemented in PyTorch [28] with timm [29] and scikit-learn [30].

Model combination

Six combination strategies were compared: feature fusion (concatenated embeddings), a greedy meta-ensemble over bases, fusion, and stacking, soft-vote, rank-mean, logistic stacking, and greedy (Caruana)

ensembling, with all weights and meta-learners fit on train + val out-of-fold predictions only, never on the frozen test.

Statistics

AUROC and Area Under the Precision-Recall Curve (PR-AUC), sensitivity at fixed specificity, and expected calibration error were computed on the frozen test; 95% confidence intervals are 2,000-resample bootstraps. Our alpha was 0.05. AUROC differences were tested with DeLong's method [22]. Operating thresholds were fixed on validation data. Inter-grader agreement was quantified on the binary open-globe target using percent agreement and Gwet's AC1, chosen over Cohen's κ because the low positive prevalence places these data in the regime where κ is unstable.

Declarations

Data availability

The curated cohort comprises publicly available web-scraped images.

Code availability

Full results and analysis code can be made available at github.com/Chuter-lab/OGI_Triage as a tagged release on acceptance.

Author contributions

B.C. conceived the study, built the data pipeline and models, performed the analysis and interpretability, and drafted and edited the manuscript. F.B.K., M.Y.K., and B.R.M. contributed to data curation, labeling and review, and submission preparation. F.M., T.L., P.C., G.J., T.M., and J.M. contributed to image screening, grading, and review. M.M.J. provided computational resources, manuscript review, editorial guidance. W.G.G. provided label adjudication and clinical guidance. All authors reviewed and approved the manuscript.

IRB/Ethics Declarations

This study used only publicly available, previously published external-eye images and did not involve interaction with human participants or access to identifiable private information. The Dartmouth Health Institutional Review Board therefore determined that this study is not research involving human subjects and that IRB review and approval were not required (STUDY02003516, PI W. G. Gensheimer, 16 June 2026).

Funding Declaration

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interests

The authors declare no competing interests.

Acknowledgements

None.

References

1. Zhou, Y. *et al.* Open globe injuries: review of evaluation, management, and surgical pearls. *Clin. Ophthalmol.* 16, 2545–2559 (2022).
2. McMaster, D. *et al.* Early versus delayed timing of primary repair after open-globe injury: a systematic review and meta-analysis. *Ophthalmology* 132, 431–441 (2025).
3. Mir, T. A. *et al.* Characteristics of open globe injuries in the United States from 2006 to 2014. *JAMA Ophthalmol.* 138, 268–275 (2020).
4. Lee, I., Davis, B., Purt, B. & DesRosiers, T. Ocular Trauma and Traumatic Brain Injury on the Battlefield: A Systematic Review After 20 Years of Fighting the Global War on Terror. *Mil. Med.* 188, 2916–2923 (2023).
5. Vlasov, A., Ryan, D. S., Ludlow, S., Weichel, E. D. & Colyer, M. H. Causes of combat ocular trauma-related blindness from Operation Iraqi Freedom and Enduring Freedom. J. Trauma Acute Care Surg. 79, S210–S215 (2015). doi:10.1097/TA.0000000000000666.
6. Ari, A. Eye injuries on the battlefields of Iraq and Afghanistan: public health implications. *Optom. St Louis Mo* 77, 329–39 (2006).
7. Weichel, E., Colyer, M., Ludlow, S., Bower, K. & Eiseman, A. Combat ocular trauma visual outcomes during operations iraqi and enduring freedom. *Ophthalmology* 115, 2235–45 (2008).
8. Dhillon, A., Ahmad, M. S. Z., Breeze, J. & Blanch, R. J. Prolonged deployed hospital care in the management of military eye injuries. *Eye* 34, 2106–2111 (2020).
9. Harvey, M. *et al.* Ocular Trauma in Operation Iraqi Freedom and Operation Enduring Freedom from 2001 to 2011: A Bayesian Network Analysis. *Ophthalmic Epidemiol.* 28, 312–321 (2021).
10. Harris, J. P. *et al.* Open-globe wounds in operation Iraqi Freedom and Operation Enduring Freedom: risk factors for poor visual outcomes and enucleation. *Acta Ophthalmol. (Copenh.)* 99, 904–908 (2021).
11. Reynolds, M. E. et al. Evaluation and treatment of ocular injuries and vision-threatening conditions in prolonged field care. J. Spec. Oper. Med. 17, 115–126 (2017).
12. Rana, V. *et al.* Combat ocular trauma in counterinsurgency operations. *Indian J. Ophthalmol.* 71, 3615–3619 (2023).
13. Gensheimer, W. G. et al. Eye Trauma: Initial Care. Joint Trauma System Clinical Practice Guideline (JTS CPG). Defense Health Agency, Joint Trauma System (2021).

14. Négrel, A. & Thylefors, B. The global impact of eye injuries. *Ophthalmic Epidemiol.* 5, 143–69 (1998).
15. Li, C. *et al.* The global incidence and disability of eye injury: an analysis from the Global Burden of Disease Study 2019. *eClinicalMedicine* 62, 102134 (2023).
16. Wang, Z. *et al.* Persistent global inequality in eye injury burden: a secondary analysis of the 2021 global burden of disease study (1990–2021). *BMJ Open* 15, e097537 (2025).
17. Ter Wei, C., Rahmat, J., Mohamad, R. & Mohd Khialdin, S. A Retrospective Analysis of Epidemiological Factors and Surgical Outcomes of Patients With Ocular Trauma at Hospital Kuala Lumpur, Malaysia. *Cureus* https://doi.org/10.7759/cureus.75941 (2024) doi:10.7759/cureus.75941.
18. Zungu, T., Mdala, S., Manda, C., Twabi, H. S. & Kayange, P. Characteristics and visual outcome of ocular trauma patients at Queen Elizabeth Central Hospital in Malawi. *PLOS ONE* 16, e0246155 (2021).
19. Beyuo, V. M., Braimah, I. Z., Abaidoo, B., Lassey, A. & Essuman, V. A. Epidemiology and factors influencing visual outcome in paediatric ocular trauma at a tertiary health institution in Ghana. Ghana Med. J. 59, 181–190 (2025).
20. Kyei, S., Kwarteng, M. A., Asare, F. A., Jemitara, M. & Mtuwa, C. N. Ocular trauma among patients attending a tertiary teaching hospital in Zimbabwe. *PLOS ONE* 18, e0292392 (2023).
21. Resnikoff, S. *et al.* Estimated number of ophthalmologists worldwide (International Council of Ophthalmology update): will we meet the needs? *Br. J. Ophthalmol.* 104, 588–592 (2019).
22. DeLong, E., DeLong, D. & Clarke-Pearson, D. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44, 837–45 (1988).
23. Qiu, J. et al. Development and validation of a multimodal multitask vision foundation model for generalist ophthalmic artificial intelligence. NEJM AI 1, AIoa2300221 (2024). https://doi.org/10.1056/AIoa2300221. Preprint at https://arxiv.org/abs/2310.04992.
24. Dosovitskiy, A., Beyer, L. & Kolesnikov, A. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. Preprint at https://arxiv.org/abs/2010.11929 (2020).
25. Oquab, M., Darcet, T. & Moutakanni, T. DINOv2: Learning Robust Visual Features without Supervision. Preprint at https://arxiv.org/abs/2304.07193 (2023).
26. Liu, Z., Mao, H. & Wu, C.-Y. A ConvNet for the 2020s. Preprint at https://arxiv.org/abs/2201.03545 (2022).
27. Tan, M. & Le, Q. V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Preprint at https://arxiv.org/abs/1905.11946 (2019).
28. Paszke, A., Gross, S. & Massa, F. PyTorch: An Imperative Style, High-Performance Deep Learning Library. Preprint at https://arxiv.org/abs/1912.01703 (2019).
29. Wightman, R. PyTorch Image Models (timm). https://doi.org/10.5281/zenodo.4414861 (2019).
30. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V. & Thirion, B. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830 (2011).

Figures



Figure 1

Representative external-eye photographs by category. License-clear external-eye photographs (open globe, indeterminate, non-open globe) shown for visual review of input-image quality and the range of real-world imagery. Figure 1 image credits (all from
Wikimedia Commons, cropped to the ocular region): open globe, "Penetrating eye injury" by Bobjgalindo (CC BY-SA 4.0), "Globe rupture" by S. Bhimji MD (CC BY 4.0), "Blunt ocular trauma - iris protrusion" by Maria Sieglinda von Nudeldorf (CC BY-SA 4.0); indeterminate, "Hyposphagma rechts Blick leicht nach unten" by Andrea Kamphuis (CC BY-SA 4.0), hyphema image from Okunuki et al., BMC Ophthalmol. 2014;14:74 (CC BY 4.0), "Hyphema - occupying half of anterior chamber of eye" by Rakesh Ahuja MD (CC BY-SA 2.5); not open globe, “Black eye as result of contusion" by Mariano-J (CC BY-SA 3.0), "Human

cornea with abrasion highlighted by fluorescein staining" by James Heilman MD (CC BY-SA 3.0), “Subconjunctival hemorrhage, small and temporal" by Iceclanl (CC BY-SA 4.0). This figure contains third-party images licensed under their own Creative Commons terms as credited above; they are not covered by the article's CC BY license.

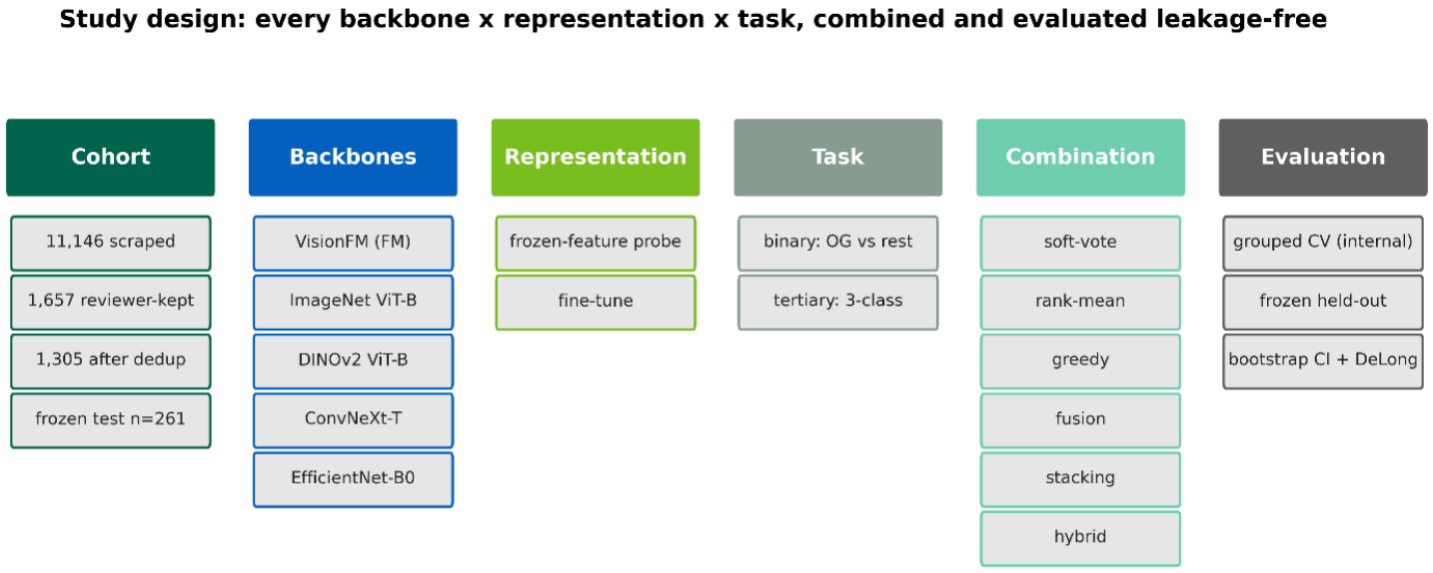


Figure 2

Study design. A single leakage-safe cohort (dedup-cluster-grouped, label-stratified) is evaluated across five backbones, two representation modes (frozen-feature linear probe and fine-tuning), two label tasks (binary open globe-vs-rest and tertiary), and six combination strategies, scored by internal grouped cross-validation and a frozen held-out test with bootstrap and DeLong inference.

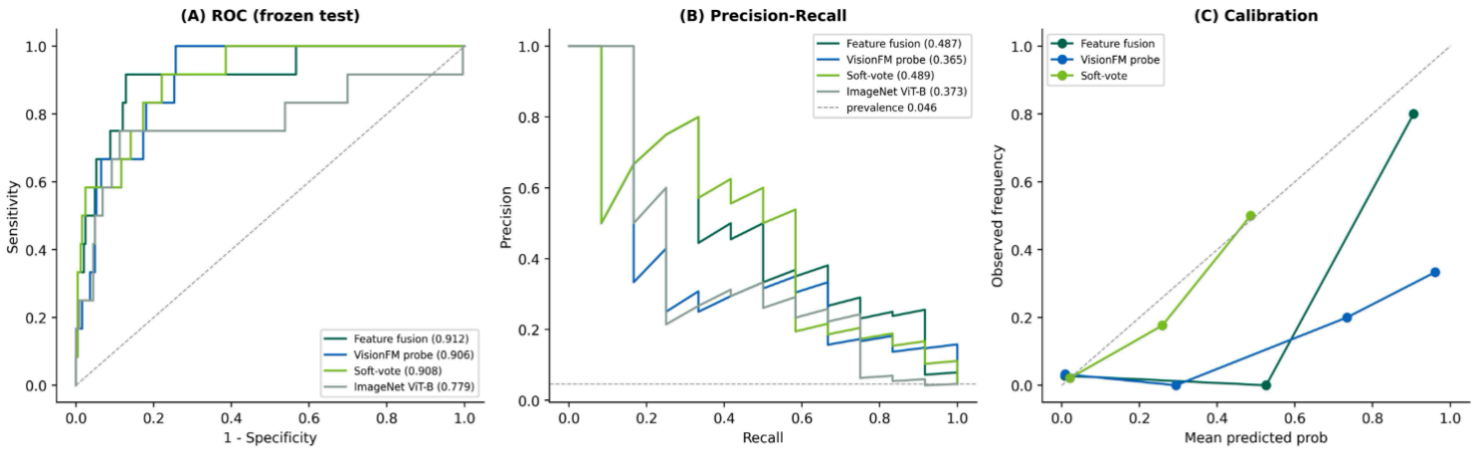


Figure 3

Discrimination and calibration on the frozen held-out test. (A) Receiver-operating-characteristic curves for the leading models (feature fusion, VisionFM probe, soft-vote ensemble, ImageNet ViT-B/16). (B)

Precision-recall curves (positive prevalence 0.046 marked as baseline). (C) Calibration (reliability) curves for the three best-calibrated models.

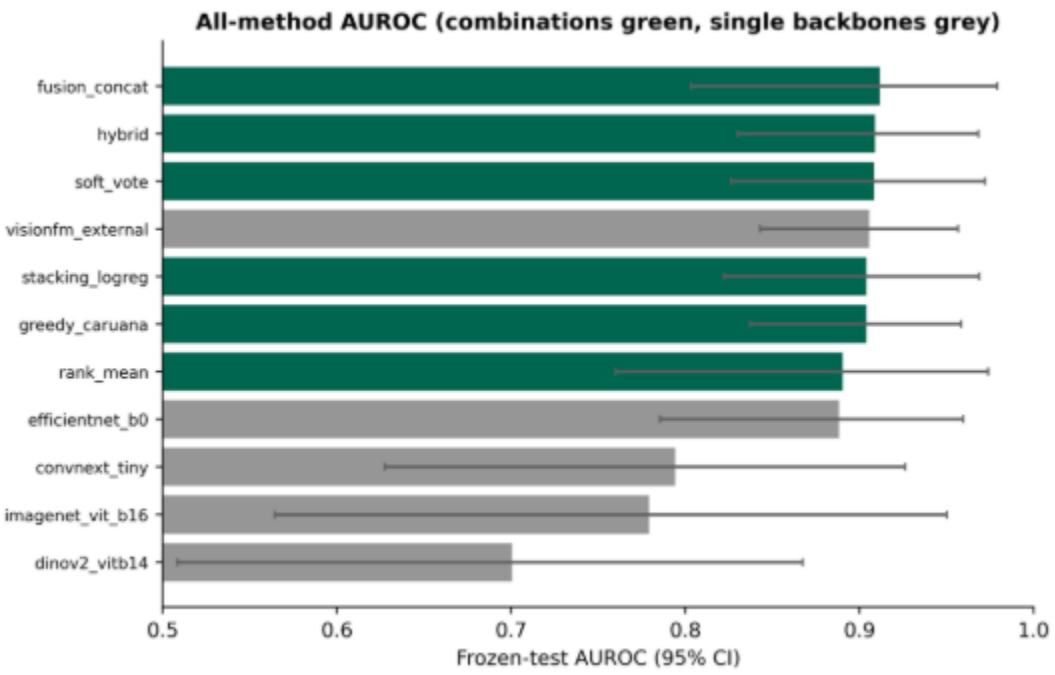


Figure 4

All-method comparison. All-method AUROC comparison with 95% bootstrap confidence intervals (combinations versus single backbones).

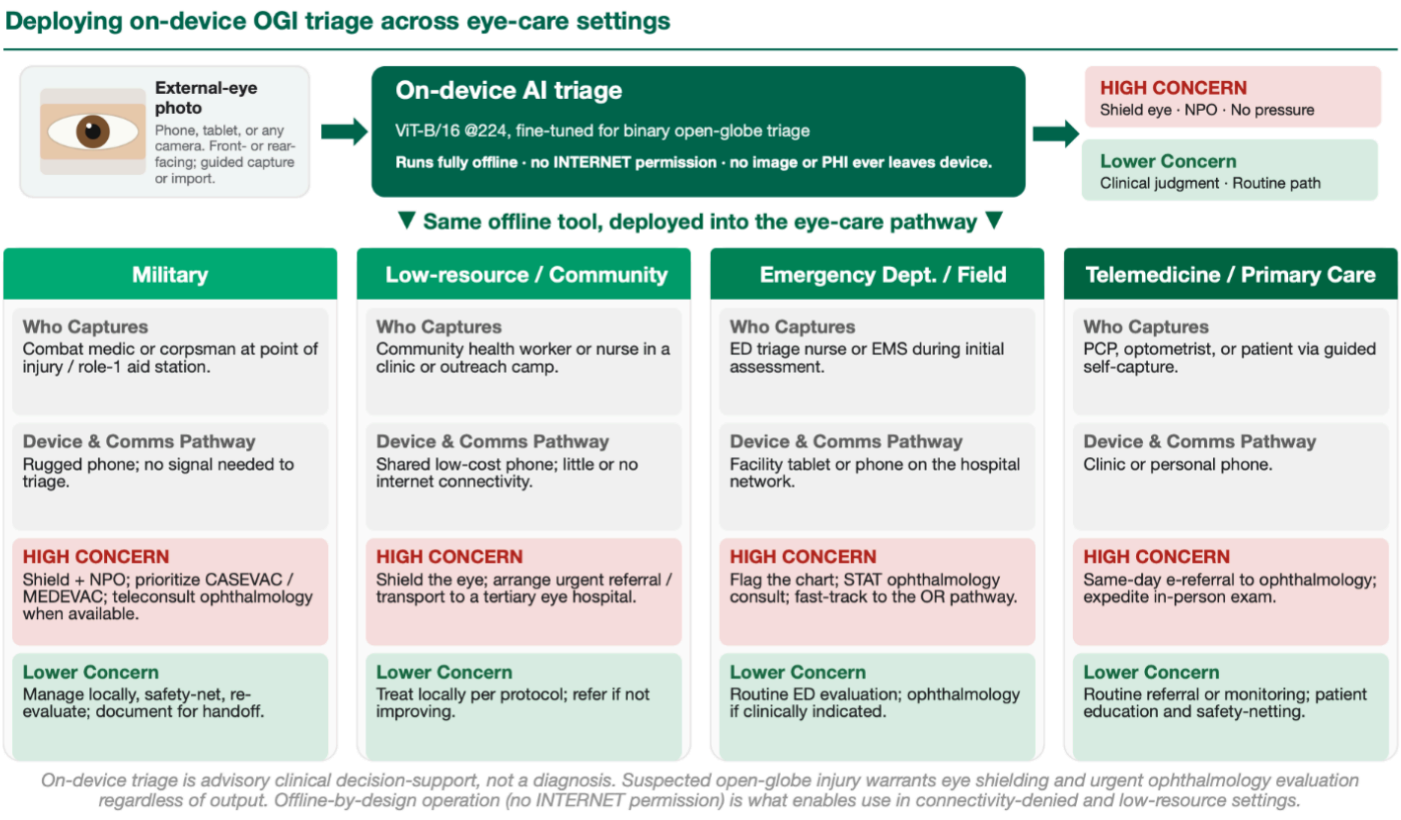


Figure 5

Deployment of on-device open globe triage across eye-care settings. The same offline model (the fine-tuned ViT-B/16, the study's top model) is deployed into four representative pathways (military role-1, low-resource community, emergency department, and telemedicine or primary care), each showing who captures the photograph, the device and connectivity pathway, and the high-concern versus lower-concern actions. Inference runs fully on-device; no image or protected health information leaves the device. History feature allows all previous images and analyses to be saved for clinician's use, and to be selectively deleted. The tool provides advisory triage decision-support, not a diagnosis.

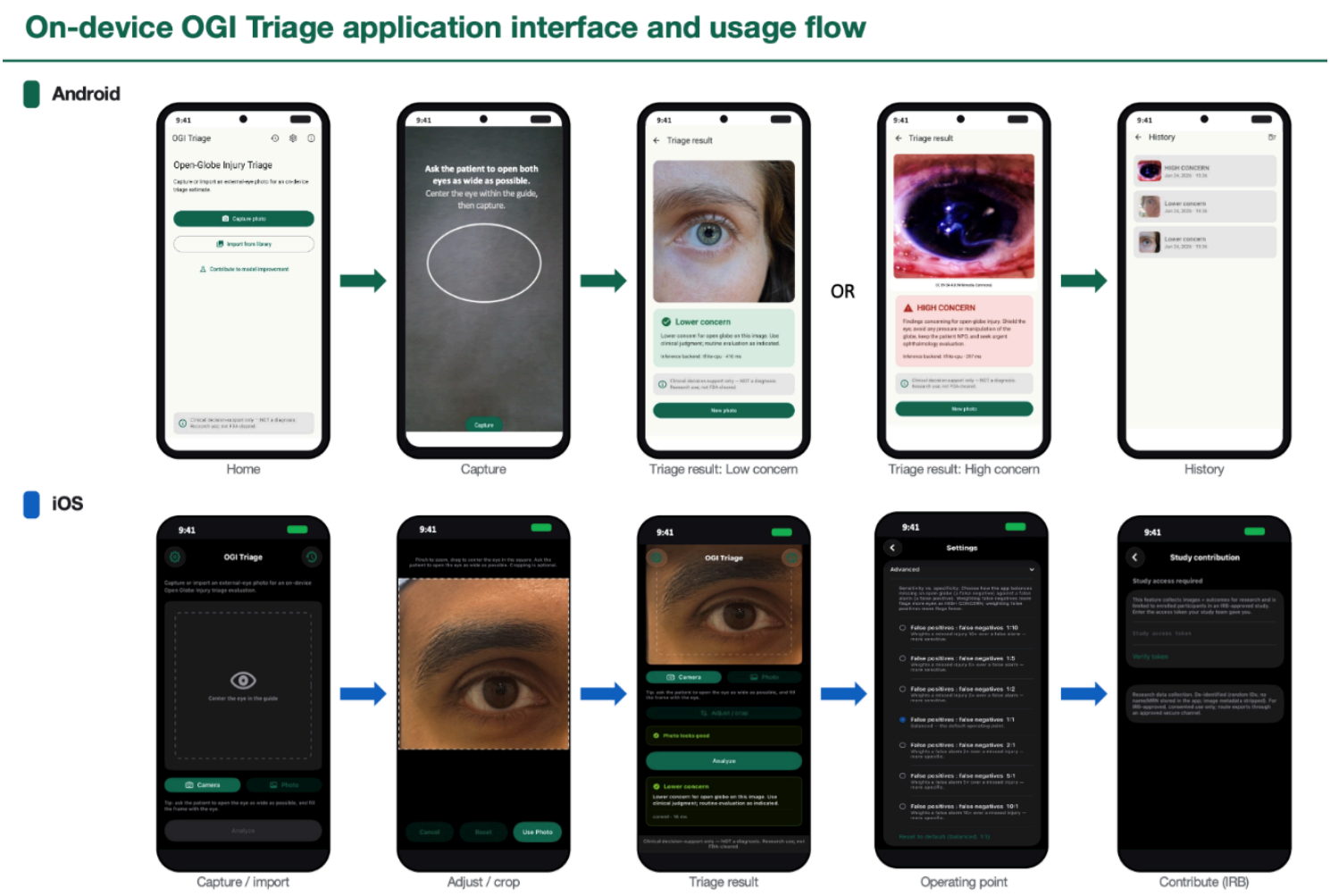


Figure 6

On-device OGI Triage application interface and usage flow. Representative Android (top) and iOS (bottom) screens from capture through triage result and study contribution, shown on real external-eye photographs. The application returns a high-concern versus lower-concern call without displaying a raw probability, and operates fully offline.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Graphicalabstract.jpg
- SupplementaryFigures.docx