Int J Ophthalmol, Vol. 19, No. 9, Sep. 18, 2026 www.ijo.cn
Tel: 8629-82245172 8629-82210956 Email: ijopress@163.com

· Review Article ·

Why aren't we using AI in eye clinics? A systematic review of barriers and solutions in AI-based fundus image diagnostics for ocular diseases

Tehmina Shehryar[1,2], *Zoha Zahid Fazal*[1], *Anum Abdul Salam*[3], *Santhoshi Varada*[1], *Fakhar Jabran*[2], *Muhammad Usman Akram*[3]

[1]Byers Eye Institute, Stanford University School of Medicine, Palo Alto 94303, CA, United States
[2]Software Engineering, Mirpur University of Science and Technology, AJK 10250, Pakistan
[3]Computer and Software Engineering, National University of Sciences and Technology, Islamabad 44000, Punjab, Pakistan
Correspondence to: Tehmina Shehryar. Stanford University School of Medicine, CA 94303, USA; Mirpur University of Science and Technology, AJK 10250, Pakistan. tehminas@stanford.edu; tehmina.se@must.edu.pk
Received: 2025-07-31 Accepted: 2026-06-08

Abstract
● Artificial intelligence (AI) has shown remarkable accuracy in the diagnosis of common ocular diseases such as diabetic retinopathy (DR), glaucoma, retinopathy of prematurity (ROP), and age-related macular degeneration (AMD), often matching or even outperforming expert clinicians. Despite these advancements, AI adoption in clinical settings remains limited due to key barriers. This systematic review evaluates 34 studies (2018–2025) highlighting AI's diagnostic performance (often >90% accuracy) while pointing out significant gaps in real-world deployment. We identify these persistent challenges through comprehensive analysis of current literature and propose actionable pathways to bridge the "last-mile gap" between research and clinical practice. This review pointed out three significant gaps in real-world deployment. These include 1) disjointed integration into clinical workflows, 2) lack of transparency in AI decision-making, and 3) poor generalizability across diverse populations. Our findings provide a framework for advancing AI implementation in ocular diagnostics to achieve equitable, scalable, and trustworthy solutions for global vision care.
● **KEYWORDS:** AI in ophthalmology; clinical adoption barriers; explainable AI; edge computing; diabetic retinopathy; glaucoma; macular degeneration

DOI:10.18240/ijo.2026.09.17

Citation: Shehryar T, Fazal ZZ, Salam AA, Varada S, Jabran F, Akram MU. Why aren't we using AI in eye clinics? A systematic review of barriers and solutions in AI-based fundus image diagnostics for ocular diseases. *Int J Ophthalmol* 2026;19(9):1809-1825

INTRODUCTION

The global burden of common ocular diseases, particularly diabetic retinopathy (DR), glaucoma, retinopathy of prematurity (ROP), and age-related macular degeneration (AMD), continues to rise due to suboptimal care. Notably, of the 2.2 billion people affected by vision impairment worldwide, half of these cases could have been prevented by timely intervention with early diagnosis[1]. Addressing this pressing issue, artificial intelligence (AI), particularly deep learning-based (CAD) systems, has revolutionized the screening and diagnosis of eye diseases and has become a gamechanger.

Recent research demonstrates that AI models can identify conditions such as DR, glaucoma, and AMD with accuracy levels comparable to or surpassing that of skilled ophthalmologist[2-4]. Despite these technological advancements, there is still limited use of AI in clinical settings. Key challenges include model generalizability across populations, lack of explainability, and disintegration of clinical workflow[5-7]. A 2025 review of the United States Food and Drug Administration-approved systems reported that although some AI tools are in active use for procedures such as DR screening, the overall adoption of AI in healthcare remains limited even in well-resourced healthcare institutions[8-10].

Additionally, widespread adaptation is further hampered by concerns regarding data biases, medicolegal liability, and lack of trust among clinicians[9]. This systematic review aims to critically examine the recent advancements, documented diagnostic results, and real-world applications of AI tools, particularly in ophthalmic care. Further, it also identifies the barriers hindering clinical adoption and proposes strategies to enhance the reliability, usability, and acceptance of AI-based diagnostics in ophthalmology.

METHODS

Search Strategy Our study follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to conduct a comprehensive literature search on Scopus, PubMed, and Web of Science databases from July 2015 through June 2025. MeSH terms and keywords were the following: "artificial intelligence", "deep learning", "machine learning" AND "fundus photography", "retinal imaging", AND "ocular disease", "diabetic retinopathy", "glaucoma", "age-related macular degeneration" AND "barriers", "challenges", AND "implementation", "integration", "clinical adoption", "workflow", "explainability", "regulatory". The search strategy was tailored to each database and refined with the help of an academic health sciences librarian.

Eligibility Criteria The inclusion criteria for our study were as follows: Population: Human subjects undergoing diagnostic evaluation for ocular disease using fundus imaging; Intervention: Use of AI-based diagnostic algorithms applied to fundus images; Outcomes: Reports on barriers to clinical implementation of AI, and proposed solutions; Studies: Original peer-reviewed articles, including clinical trials and observational studies in English.

The exclusion criteria set for this review were as follows: Study types: reviews, opinion pieces, editorials; Modalities: studies that used optical coherence tomography (OCT) or slit-lamp; Research strategy: not addressing clinical deployment challenges.

Study Selection After duplicate removal, two independent reviewers screened titles and abstracts for relevance. Full-text review was performed for articles that met inclusion criteria or could not be excluded based on abstract alone. Discrepancies were resolved through consensus or adjudication by a third reviewer. The selection process was documented using a PRISMA flow diagram.

Data Extraction A standardized data extraction form was developed and pilot-tested. Extracted data included: Publication year, country, and study design; AI model type (*e.g.*, ensemble models) and validation strategy; Disease target(s) (*e.g.*, DR, AMD, glaucoma); Implementation context (*e.g.*, primary care, tele-ophthalmology, hospital setting); Reported barriers (*e.g.*, integration, accuracy, cost, liability, explainability, user trust); Proposed solutions [*e.g.*, explainable AI (XAI), clinical validation, workflow redesign, regulatory guidance].

Risk of Bias and Quality Assessment The Mixed Methods Appraisal Tool (MMAT, 2018) was used to assess study quality across various designs. Two reviewers independently rated the quality of included studies, with disagreements resolved by consensus. No studies were excluded based on quality score, but findings were interpreted in the context of study rigor.

Synthesis of Results We employed a thematic analysis approach to categorize reported barriers and proposed solutions. Barriers were grouped into higher-order themes including technical limitations, workflow integration challenges, regulatory/legal hurdles, trust and explainability issues, and economic/resource constraints. A narrative synthesis was used to summarize findings across studies, supported by frequency counts and illustrative quotes or examples where available.

RESULTS

Study Selection The database search initially retrieved 2347 records. After removing 764 duplicates, 1583 titles and abstracts were screened. Of these, 215 full-text articles were assessed for eligibility. Ultimately, 34 studies met all inclusion criteria and were included in the final synthesis.

Study Characteristics The included studies were published between 2016 and 2025, with an increasing trend after 2019. The majority were conducted in high-income countries, with the most studies originating from the United States (n=12), China (n=7), India (n=5), and the United Kingdom (n=3). Study designs comprised of technical validation or model development studies (n=14), implementation case studies or feasibility pilots (n=9), mixed-methods evaluations (n=6), and qualitative studies, clinician/user surveys (n=5). AI models focused primarily on convolutional neural networks (CNNs), with several studies (n=8) incorporating explainability techniques such as attention maps or saliency overlays. The most frequently studied diseases were DR (n=25), glaucoma (n=10), and AMD (n=8). Some studies evaluated multiple diseases.

Barriers to Clinical Implementation and Their Proposed Solutions The literature demonstrates that although automated diagnosis of retinal diseases is at a mature level as summarized in Table 1[10-38], unfortunately, its application in clinical practice is minimum due to the following main reasons.

Barrier 1: disjointed integration into clinical workflows and the complexity, cost, and resource demands of AI-based CAD systems limit their adoption

1) Rationale Smartphone fundoscopy has emerged as a user-friendly and cost-effective alternative to conventional digital cameras for capturing fundus images. It is a potential technique for retinal imaging because of its connection, cost, usability, and accessibility. Table 2[39-42] summarizes the recent use of AI-based retinal screening algorithms using smartphone fundoscopy. This makes it especially valuable for underserved and resource-constrained settings, where access to specialist care and advanced retinal imaging is limited, and screening is often needed in community clinics or outreach settings. However, in many real-world clinical settings, AI tools still remain difficult to adopt because they often require extra steps

Table 1 Performance evaluation and accuracy of AI-based retinal screening algorithms using fundus images

Author	Year	Retinal diseases	AI technique	Mobile app	GPU	Dataset	Accuracy (%)
Skevas *et al*[10]	2022	DR+AMD	RetCAD (cloud-based AI on fundus photos)	No	Yes	630 patients; 1245 eyes; 3609 fundus photos	96.1
Ashikur *et al*[11]	2020	DR	Inception-v4	No	Yes	53576	91.5
Sudha *et al*[12]	2023	DR	Cascaded Rotation Forest	No	Yes	560	98.0
Madi *et al*[13]	2026	DR	URNet/ViT/CNN Ensembles (EyeArt)	No	Yes	100000+	98.5
Kour and Saini[14]	2026	DR	Attention-guided CNN	No	Yes	APTOS 2019 (3662)	99.8
Al-Shalabi *et al*[15]	2026	DR	Hybrid machine learning	No	Yes	Messidor (1200), IDRiD (516)	99.1
Cao *et al*[16]	2019	DR	GLCM texture features+Naïve Bayes	No	No	1000 fundus images (298 DR, 702 non-DR)	93.5
Sundaram *et al*[17]	2023	DR	Inception-v4	No	Yes	8739	94.0
Velcani[18]	2026	DR/cognitive decline	Deep learning biomarker extraction	No	Yes	Clinical cohort (2280)	73.0
Singh *et al*[19]	2022	Glaucoma	BA, BCS, PSO	No	Yes	ORIGA (650) REFUGE (1200)	99.0
Shoukat *et al*[20]	2021	Glaucoma	EfficientNetB7	No	Yes	G1020 (1020)	99.3
Zedan *et al*[21]	2023	Glaucoma	Random forest	No	Yes	Drishti-GS (101)	99.0
Raja *et al*[22]	2021	Glaucoma	SVM	No	Yes	175	92.0
Kumar *et al*[23]	2026	Glaucoma	GlaucoNet Ensemble (MobileNetV2+InceptionV3+ResNet50)	No	Yes	ORIGA (650); ACRIMA (705)	97.0
Gandhi *et al*[24]	2026	Glaucoma	Multi-modal vision transformer (Fundus+OCT)	No	Yes	DRISHTI-GS; ORIGA; RIGA; DRIONS-DB; GlaS	94.5
Khalid *et al*[25]	2021	AMD	SVM	No	Yes	462	97.5
Boukadida *et al*[26]	2022	AMD	VGG16 neural network	No	Yes	395	98.3
Xu *et al*[27]	2021	AMD	ResNet-50	No	Yes	143	92.0
Raja Sankari *et al*[28]	2023	ROP	MultiResUNet	No	Yes	4000	94.0
Tong *et al*[29]	2020	ROP	Faster-RCNN	No	Yes	36231	95.7
Chen *et al*[30]	2021	ROP	CNN	No	Yes	10992	98.0
Mowla *et al*[31]	2025	ROP	LightEyeNet (Attention DenseNet)	Yes	No	eHealth multi-class (1840)	96.3
Zhao *et al*[32]	2026	ROP	Multimodal AI data evaluation	No	Yes	G-ROP registry (7483)	96.0
Mohiy *et al*[33]	2026	ROP	ROPDeepX (ResNet50+EfficientNet)	No	Yes	FARFUM/RIDIRP (2430)	96.9
Vahidmoghadam *et al*[34]	2026	ROP	EfficientNetB4+vessel segmentation (U-Net++)	No	Yes	RIDIRP (6004) [Plus]; RIDIRP (880) [Stage]	99.6
Oulhadj *et al*[35]	2024	DR	ViT+Modified capsule network	No	Yes	APTOS (3662)	88.2
Tampa *et al*[36]	2025	Glaucoma	Modified VGGNet19+transfer learning	No	Yes	ORIGA/ORIGA-light (1300)	98.8
Chaurasia *et al*[37]	2025	Glaucoma	VGG19_bn+transfer learning	No	Yes	20 public datasets (18468)	98.8
Ikram and Imran[38]	2025	DR	ResViT FusionNet (ResNet50+ViT)	No	Yes	APTOS 2019 (5590)	93.0

ACRIMA: Automated classification of retinal images for glaucoma assessment; AI: Artificial intelligence; AMD: Age-related macular degeneration; APTOS: Asia Pacific tele-ophthalmology society blindness detection dataset; BA: Bat algorithm; BCS: Binary cuckoo search; CNN: Convolutional neural network; DR: Diabetic retinopathy; DRIONS-DB: Digital retinal images for optic nerve segmentation database; EfficientNet: Efficient neural network; EyeArt: EyeArt automated diabetic retinopathy screening system; FARFUM: Fundus analysis for retinopathy of prematurity and follow-up management dataset; GLCM: Gray-level co-occurrence matrix; GPU: Graphics processing unit; OCT: Optical coherence tomography; ORIGA: Online retinal fundus image database for glaucoma analysis; ORIGA-light: Online retinal fundus image database for glaucoma analysis-light version; PSO: Particle swarm optimization; ReTCAD: Retinal computer-aided diagnosis system; REFUGE: Retinal fundus glaucoma challenge dataset; ResNet: Residual neural network; RIDIRP: Retinal image dataset for infant retinopathy of prematurity; ROP: Retinopathy of prematurity; ROPNet: Retinopathy of prematurity network; SVM: Support vector machine; U-Net: U-shaped convolutional neural network; ViT: Vision transformer; VGG: Visual geometry group network.

Table 2 AI-based retinal screening algorithms using smartphone fundoscopy

Author	Year	Retinal diseases	Imaging modality	AI technique	Mobile app	GPU	Dataset	Performance (%)
Mrad *et al*[39]	2022	Glaucoma	Smartphone fundus cameras	SVM	Yes	No	24	Acc. 100
Rajalakshmi *et al*[40]	2018	DR	Smartphone fundus cameras	EyeArtTM	No	Yes	301	Sen. 95.8, Spec. 80.2
Hacisoftaoglu *et al*[41]	2020	DR	Synthetic smartphone fundus cameras	ResNet50	No	Yes	45000	Acc. 99.1
Wroblewski *et al*[42]	2025	DR	Remidio FOP smartphone camera	Medios (offline)&EyeArt (online)	No	Yes	2130	Medios: Sen. 94, Spec. 94, EyeArt: Sen. 94, Spec. 86

Acc.: Accuracy; AI: Artificial intelligence; DR: Diabetic retinopathy; FOP: Fundus-on-phone; GPU: Graphics processing unit; Sen.: Sensitivity; Spec.: Specificity; SVM: Support vector machine; ResNet: Residual neural network; EyeArt: EyeArt automated diabetic retinopathy screening system.

Table 3 Algorithm optimization techniques for edge AI

Authors	Year	Optimization technique	Methodology	Key findings/results
Tan and Wang[43]	2021	Model compression	Sparse regularization, iterative pruning, clustering-based quantization	Reduced model size without significant performance loss
Qin *et al*[44]	2018	Quantization&pruning	Evaluated 11 NN (NVIDIA Jetson Tx2) architectures with quantization/pruning	Reduced storage, faster inference, lower energy use
Matsubara *et al*[45]	2020	Knowledge distillation	Early-layer in-network compression *via* teacher-student training	High compression while maintaining accuracy
You and Tang[46]	2021	Metaheuristic optimization	Multi-user MEC task offloading with penalty functions	PSO outperformed GA/SA in scalability
Hosseinzadeh *et al*[47]	2021	Butterfly optimization	Task prioritization+DVFS scheduling in MEC	Lower energy/data access overheads
Prihozhy *et al*[48]	2000	Parallelization	VHDL-based async network parallelization	Improved concurrency in distributed systems
Yoo[49]	2018	Heterogeneous CPU design	Hybrid high-performance+energy-efficient (ARM) cores	Dynamic power-performance scaling
Jang *et al*[50]	2020	On-device training	DRL policy compression *via* knowledge distillation	Near-cloud performance with shorter training
Kaneko *et al*[51]	2019	Low-power backpropagation	Fixed-point+ternarized gradients	100× lower power *vs* 16-bit quantization
Gurnani and Kaur[52]	2026	Clinical translation validation	Multi-site edge evaluation of automated screening networks	Bridges the gap between sandbox innovation and point-of-care patient triage
Altayeb *et al*[53]	2026	Explainable multi-staging	Multi-stage Swin and Vision Transformers with attention mappings	Achieves Acc.=94.15% across varied multi-disease ocular datasets
Baddur and Sangaralingam[54]	2026	Decentralized optimization	Parallel convolutional LeNet (PC-LeNet) with firefly-assisted loops	Improves diagnostic stability and speeds up multi-center convergence
Cacciatore *et al*[55]	2026	Contrastive pretraining	Privacy-preserving federated contrastive weight alignment layers	Ensures secure cross-domain generalization on entirely unseen registries

Acc.: Accuracy; AI: Artificial intelligence; ARM: Advanced RISC machines; CPU: Central processing unit; DRL: Deep reinforcement learning; DVFS: Dynamic voltage and frequency scaling; GA: Genetic algorithm; MEC: Mobile edge computing; NN: Neural network; PSO: Particle swarm optimization; SA: Simulated annealing; VHDL: VHSIC hardware description language.

(separate software, manual upload/export of images, extra documentation), expensive hardware, and do not fit naturally into the day-to-day workflow of clinics.

2) Proposed solution The complexity of the CAD solutions involving multiple machines, time-consuming processes and high costs may also deter the practitioners from adopting new techniques. To deal with all such complexities, a potential solution could be the development of an innovative edge device-based diagnostic system that utilizes smartphone fundoscopy to enable rapid and accessible healthcare assessment for these prevalent eye diseases. Table 3[43-55] summarizes the algorithm optimization techniques for edge AI. This innovative solution can enable the early diagnosis of potential patients who have a higher likelihood of certain ocular diseases, facilitating timely referral for detailed analysis and appropriate treatment. The key advantage of this edge device-based diagnostic system is twofold. Primarily, it will significantly reduce the cost and complexity associated with traditional fundus cameras and expensive equipment such as graphics processing unit (GPU), making inexpensive eye screening and diagnosis easily accessible in both rural and urban areas. With on-device processing instead of relying on a dedicated GPU in the cloud, we can overcome limitations associated with latency, privacy concerns, and limited connectivity. Second, it will increase the adoption of automated ocular disease diagnostic systems in clinical setups due to its seamless integration with existing clinical workflow. Integrating ocular disease diagnostic systems into mobile phones allows for easy and convenient access to diagnostic tools and data analysis capabilities. The proposed system can be a path-breaking innovation which could revolutionize the way eye diseases are being detected by giving an affordable, portable, and hassle-free alternative involving mobile phones.

To make this seamless integration practical, we envision that such systems can be embedded into clinical care at three levels: 1) Point-of-care screening (primary care/diabetes clinics): image capture (fundus camera or smartphone adapter)–automated image quality check–AI triage (refer/urgent/non-urgent)–referral pathway; 2) Ophthalmology clinics: AI can act as a "second reader" integrated with routine imaging review systems and/or picture archiving and communication system (PACS)/electronic health record (EHR) to pre-screen images, highlight suspected lesions/regions, and support patient prioritization, while the clinician retains the final decision; 3) Tele-ophthalmology/outreach screening: remote image acquisition to AI pre-read (online or offline) to asynchronous ophthalmologist review for positive/uncertain cases; 4) Importantly, this workflow is also aligned with global vision care needs: smartphone-based edge AI can support low-cost screening, offline or low-connectivity operation, rapid triage at the point of care, and timely referral of high-risk patients—helping extend services to rural and underserved populations where ophthalmologists are scarce.

For real-world adoption, it is also important that these tools require minimal additional clicks/time, provide clear referral thresholds, and support documentation (*e.g.*, storing results and

Table 4 Empirical evidence on transparency and adoption

Authors	Year	Key findings	Domain
Gurnani and Kaur[52]	2026	Clinical transparency and feature-driven attention overlays are essential requirements to close the translation gap for AI screening tools	Clinical implementation
Altayeb *et al*[53]	2026	A multi-stage Swin Transformer+Vision Transformer framework achieves 94.2% classification accuracy and provides explainable attention mapping layers	Computer vision
Ribeiro *et al*[56]	2016	Users trust on AI models increases when provided with interpretable explanations	General AI
Wachter *et al*[57]	2017	Legal frameworks (*e.g.*, GDPR) needs "right to explanation" for automated decisions	Law & policy
Adeniran *et al*[58]	2024	XAI techniques enhances adoption in healthcare and finance	Healthcare/finance
Burrell[59]	2016	Black-box AI causes user skepticism due to inability to audit decisions	Social implications

AI: Artificial intelligence; GDPR: General data protection regulation; XAI: Explainable artificial intelligence.

maintaining audit trails) so clinics can use them confidently in routine practice.

Barrier 2: lack of transparency in outputs and results of AI systems also lead to skepticism and limited adoption

1) Rationale Widespread adaptation of AI model is constrained due to black box decision making processes contributing to lack of trust and skepticism among stakeholders. Transparency in AI is the ability of users to comprehend the decision-making process, understand how decisions are made, what data is used, and potential biases in system. Notable studies on AI transparency and impact on adoption is summarized in Table 4[52-53,56-59]. In real clinical practice, transparency is not only about where the model looked in a fundus image; clinicians also want to understand how the model was trained and validated, what its intended use is (screening/triage vs diagnosis), and where it may fail. Similarly, outputs must be presented in a clinically meaningful way (*e.g.*, confidence or referral rationale) so that clinicians can safely act on the AI recommendation. Key challenges include: 1) Decreased trust: if user can not verify the reason behind outputs, they are less likely to trust AI systems; 2) Ethical and legal concerns: in high-stake fields like healthcare, lack of explain-ability raises questions regarding accountability; 3) Limited adoption: organizations are hesitant to use AI models that are not interpretable.

2) Proposed solution By integrating technical approaches, such as AI techniques with regulatory frameworks and human-in-the-loop (HITL) oversight, a multi faced strategy could be an effective way to improve transparency. To address the problem of AI opacity, researchers and practitioners have put forth XAI Techniques which provide several solutions: Local interpretable model-agnostic explanations (LIME): LIME[56] explains the image classification predictions by emphasizing the most significant areas (*e.g.*, superpixels). It modifies the input image and trains a basic model such as linear regression to approximate the behavior of the complex model. At the end, it shows a saliency map highlighting the important region contributing to the prediction. Figure 1 exemplifies prediction of LIME application, by marking most influential areas.

SHapley Additive exPlanations (SHAP): The SHAP[60] approach uses cooperative game theory to identify features contributing to decision. It calculates the Shapley values by analyzing every possible feature contribution. Importance score is given to the regions that satisfy fairness accuracy consistency. It provides both local and global explanations.

Saliency and Heatmap-Based Explanations: Visual tools like heat maps[61] can highlight the region of an image that most influenced a models prediction. They are generated using techniques such as Gradient-weighted class activation mapping (Grad-CAM), saliency maps, and activation maximization. These visualizations are frequently used to determine which anatomical areas are involved and responsible for decision making. Figure 2 demonstrates an example of visualization techniques[18], showing ground-truth lesions (last column) alongside model predictions across different stride levels (columns), illustrating interpretable attention alignment with clinical annotations.

Process transparency: Along with XAI, AI systems should also provide clear documentation of training data provenance (population and device diversity), labeling/grading protocol, preprocessing steps, validation strategy (internal *vs* external), intended use, and known limitations/failure modes. In practice, structured reporting tools such as model cards can help communicate these details to clinicians and regulators[62]. In high-stakes applications like healthcare, transparency should not rely only on saliency maps or visual explanations. Practical steps include providing standardized documentation of the model's intended use, training data characteristics, validation setting, and known limitations (*e.g.*, using model cards[63]), maintaining auditability through clear documentation and audit logs (model version, date/time, output), and incorporating clinician-in-the-loop review for uncertain or borderline cases. In addition, presenting outputs with calibrated confidence/risk scores and clear referral rationale can help clinicians interpret results safely and make accountable decisions in routine practice.

Clinical transparency: For ocular disease diagnosis and screening, transparency also means presenting outputs in an

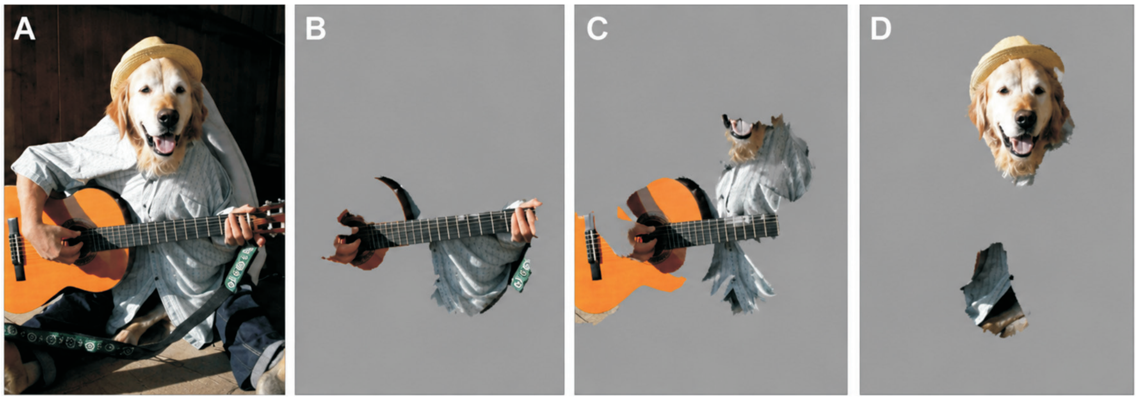

Figure 1 LIME explanation highlights influential image regions for an Inception network's top predictions A: Original image depicting a dog wearing a hat and shirt while playing an acoustic guitar; B: Visual explanation for the prediction "electric guitar", highlighting image regions that most strongly contributed to the model's classification, primarily the guitar neck, fretboard, strings, and the hand position; C: Visual explanation for the prediction "acoustic guitar", emphasizing the guitar body, sound hole, strings, and adjacent contextual features that support recognition of an acoustic guitar; D: Visual explanation for the prediction "Labrador", illustrating that the model primarily focuses on the dog's facial features, head shape, ears, and other discriminative anatomical characteristics when identifying the breed. Highlighted regions indicate the image areas that contributed most to each prediction, demonstrating that the model attends to distinct features depending on the classification task. LIME: Local interpretable model-agnostic explanations.

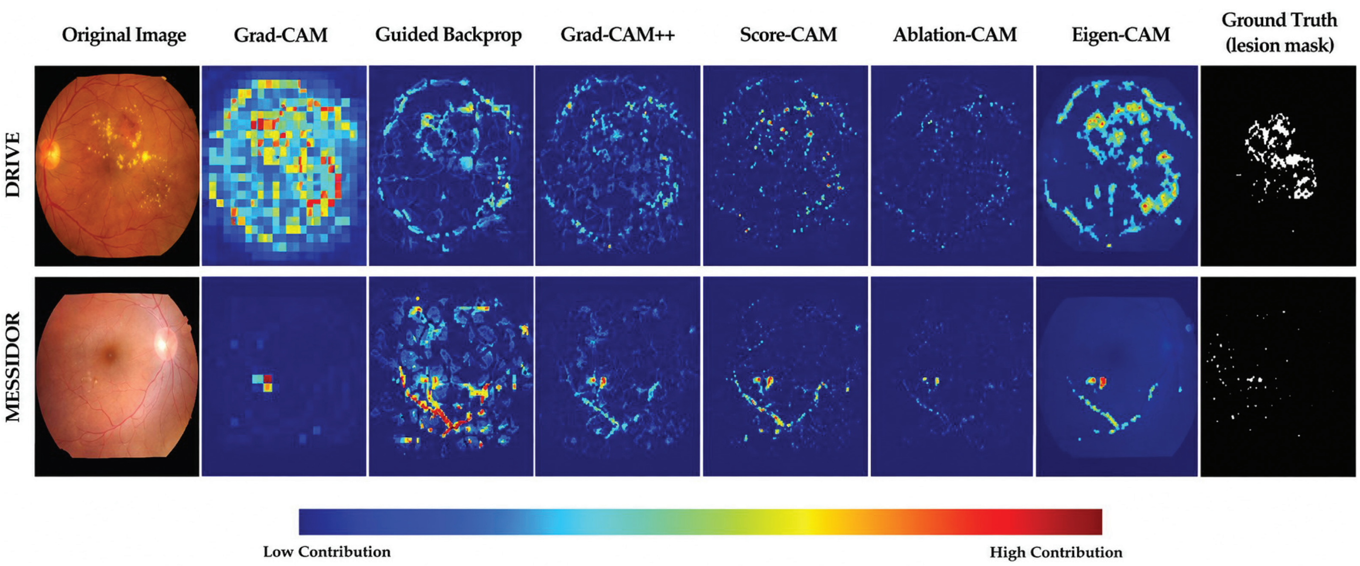

Figure 2 Focused attention Ground truth lesions and aggregated results are highlighted in the last columns, while interpretable predictions are displayed across tokenization levels (columns)[18]. Grad-CAM: Gradient-weighted class activation mapping.

actionable format: clear referral categories (*e.g.*, no referral/routine/urgent), calibrated confidence or risk scores, and a brief rationale supported by the explanation map where appropriate. Maintaining documentation/audit logs (model version, time, output) and using HITL review for uncertain cases can further strengthen accountability and trust in high-stakes clinical settings[4-5].

Barrier 3: lack of generalization in AI models often stems from biases in training data, which causes misdiagnosis, delayed treatment, and worse health outcomes for patients from under-representative groups

1) Rationale AI has gained clinical acumen in ophthalmology, particularly when it comes to identifying diseases through medical imaging. However, this success is a lopsided achievement, since these AI systems come with significant limitations that require urgent attention. A growing concern is that many AI models are trained on datasets heavily skewed toward specific demographic groups. Such models frequently fail to perform accurately when used on patients from an unrepresented population. This lack of generalization is primarily due to biases in training data, which causes misdiagnosis, delayed treatment, and worse health outcomes for these groups. It will not only limit the global usefulness of AI tools but also deepen existing healthcare disparities.

In real clinical settings, poor generalizability can have direct consequences. When model performance drops in specific

Table 5 Solutions to improve generalization in AI for ophthalmology

Solution	Description
Decentralized optimization[54]	Implement decentralized frameworks optimized *via* multi-site convergence loops (*e.g.*, Firefly-assisted loops) to improve model stability across clinical centers
Contrastive pretraining[55]	Apply privacy-preserving federated contrastive pretraining and weight-alignment to support secure cross-domain generalization on unseen diagnostic registries
Diverse and representative datasets[64]	Curate datasets with varied ethnicity, age, disease stages, and device sources (*e.g.*, fundus, OCT from multiple clinics and regions)
Fairness-aware training[65]	Incorporate fairness-aware loss functions or rebalancing methods to reduce bias in retinal disease prediction among different population groups
Domain adaptation and transfer learning[66]	Use models pretrained on large datasets (*e.g.*, EyePACS) and fine-tune on local datasets to adapt to new populations and devices
Data augmentation for minority classes[67]	Use GANs, flipping, cropping, and contrast manipulation to synthetically boost under-represented retinal pathology samples
Subgroup performance evaluation[68]	Stratify model evaluation by ethnicity, age, device, or geography to identify disparities and performance gaps
Federated learning[69]	Collaborate across eye hospitals to train decentralized models that generalize without compromising patient privacy
Expert-in-the-loop feedback[70]	Engage ophthalmologists in reviewing predictions during model training or deployment to ensure clinical accuracy

AI: Artificial intelligence; EyePACS: Eye picture archive communication system; GAN: Generative adversarial network; OCT: Optical coherence tomography.

populations or on images from different devices/cameras, it can increase false negatives (missed disease and delayed referral/treatment) and false positives (unnecessary referrals, additional testing, higher costs, and increased workload for clinics). Over time, these errors can reduce clinician trust in AI systems and may widen existing health inequities if underrepresented groups consistently receive less accurate AI-supported care[5].

2) Proposed solution Table 5[54-55,64-70] summarizes potential solutions to improve generalization in AI models due to biased training data. Collectively, these practices are essential for developing AI systems that are both equitable and clinically reliable. To directly address dataset diversity, concrete steps include building multi-center datasets with demographic and device diversity, performing subgroup-based evaluation (age/ethnicity/device/geography), and using approaches such as fairness-aware training, domain adaptation, data augmentation for minority classes, federated learning, and expert-in-the-loop feedback to continuously improve performance across populations. It is emphasized that these strategies should be paired with external validation on independent datasets and transparent reporting of subgroup performance before clinical deployment.

Risk of bias Using the MMAT tool, 22 studies were rated as moderate to high quality, while 12 studies showed notable limitations, such as limited external validation, self-reported metrics, or incomplete outcome reporting. Technical validation studies tended to have higher risk of bias compared to real-world implementation or qualitative research.

DISCUSSION

Performance Evaluation and Accuracy of AI-Based Retinal Screening Algorithms Using Fundus Image In this section, we delve into the extensive body of research that focuses on the performance evaluation and accuracy assessment of AI-based retinal screening algorithms using fundus images. Mainly, AI methods for diagnosing ocular diseases from fundus images are broken down into a few simple groups: 1) end-to-end deep learning (DL) models that directly classifies the complete image (often CNNs or transformers); 2) severity classification models that assign stages (*e.g.*, DR levels); 3) segmentation and feature analysis approaches that first segment region of interest (ROI; structures/lesions) and then compute clinical bio markers such as cup-to-disc ratio (CDR) for glaucoma; 4) hybrid systems that combine extracted features with classical machine learning [ML; *e.g.*, support vector machine (SVM)/random forest] for diagnosis of disease.

Skevas *et al*[10] prospectively evaluated RetCAD, a cloud enabled AI system for simultaneous screening and grading of AMD and DR from color fundus photographs in a real clinical workflow. The study included 630 patients (1245 eyes; 3609 fundus photos) and compared AI outputs with expert grading (with adjudication for disagreements). RetCAD achieved strong performance with Az=0.964 (AMD) and Az=0.961 (DR) on image level, supporting its potential for cost effective multi-disease screening. Cao *et al*[16] proposed an automated DR screening approach using handcrafted texture features extracted from fundus images *via* the Gray Level Co-occurrence Matrix and classification with a Naïve Bayes model. The study used 1000 fundus images from diabetic patients (298 DR) and reported strong validation performance [area under the curve (AUC)=0.938, sensitivity 0.949, specificity 0.928] with 93.5% average accuracy under 10-fold cross-validation. In another recent retrospective study, a cohort of 3285 patients yielded a total of 8739 retinal fundus images[56]. A multiple enhanced Inception-v4 ensembling technique was created for the purpose of identifying DR and diabetic macular edema (DME). This work demonstrated that the deep ensemble model has strong resilience and generalization, great performance in identifying DR with 94% accuracy, and the ability to support and grow DR/DME screening programs. Ashikur *et al*[11] used Kaggle public dataset for DR grading, including 53 576 fundus photos in the test set, 28 101 in the training set, and 7025 in the validation set. Totally 4192 photos are chosen at random for lesion annotation. The categorization method is based on the Inception V3 framework. With 896×896 resolution photos as input for severe DR, the method's sensitivity, specificity, harmonic mean, and AUC are 0.925, 0.907, 0.916, and 0.968, respectively. The novel background and foreground superpixel segmentation method is proposed by the CNN-based automated detection of DR, and the primary classification of fundus image features was carried out using hybrid classifiers such as the K-Nearest Neighbour, Support Vector Machine, and cascaded rotation forest (CRF) classifiers[12]. The proposed method performs exceptionally well and attained 98% total accuracy. CRF classifier is the most accurate of them. The landscape of automated DR diagnostic frameworks has shifted drastically toward DL, ensemble models, and advanced feature extractions in 2026. This systematic review evaluates the evolution of AI models, specifically CNNs, for early DR identification across massive public registries comprising over 100 000+ public retinal fundus images. The findings highlight that specialized frameworks like EyeArt and Google AI achieve expert-level diagnostic metrics, with EyeArt matching human retinal specialists by identifying 98.5% of vision-threatening cases[13]. Expanding upon structural feature maps, recent architectures integrate explainability components to support clinical trust. An explainable, attention-guided DL framework utilizes a custom CNN feature extractor to isolate localized pathology and microaneurysms directly from retinal images using the Asia Pacific Tele-Ophthalmology Society Blindness Detection Dataset (APTOS) 2019 dataset (3662). Validated across these diverse data repositories, this attention-driven network delivers a diagnostic accuracy of 99.80% while establishing clear clinical feature mapping[14]. Parallel to image-only models, modern frameworks seek to blend demographic datasets with ocular imagery to optimize predictive reliability. A hybrid ML strategy extracts a high-dimensional feature vector containing 20 distinct morphological and textural traits from patient retinal profiles. By fusing these visual variables with structured tabular data from a combined pool incorporating the Messidor dataset (1200) and the IDRiD dataset (516), this dependable screening system maximizes classification precision in complex early-stage variations[15]. Finally, the scope of automated eye screenings has extended to systemic diagnostics through oculomics. A DL biomarker extraction model was deployed to evaluate ultra-widefield fundus photographs from insulin-dependent type 2 diabetic patients using a specialized clinical cohort (2280). The system successfully isolates hidden retinal microvascular configurations to accurately forecast localized global cognitive impairment risks directly at the point of care.

A novel two-layered automated glaucoma detection technique based on particle swarm optimization (PSO), binary cuckoo search, and bat algorithm have been proposed by Singh *et al*[19]. Separate analyses have been done on the performances of these three ML classifiers, which when fed five (single and two-layered) techniques, can provide the most accurate subsets of reduced features. To validate the suggested technique, benchmark datasets ORIGA and REFUGE are combined with other publicly accessible datasets. With these techniques, an accuracy of up to 98.95% is possible. Similarly, in another research study, for the purpose of evaluating the suggested technique, Shoukat *et al*[20] employed three datasets: G1020, RIM-ONE, and REFUGE. For the classification, three pre-trained CNN architectures such as VGGl9, ResNet50, and EfficientNetB7 are employed, and their performance is contrasted using various performance metrics. On the G1020 dataset, the EfficieientNetB7 architecture produces the best classification results with accuracy of 99.2%, sensitivity of 98%, and specificity of 97%. A glaucoma identification method is presented by Zedan *et al*[21] using fundus pictures to estimate CDR. To identify the existence of glaucoma, the size of the optic disc and optic cup are used. The glaucomatous images have been classified using a random forest classifier based on the CDR values after the cup and disc have been extracted using image processing techniques. Deformable U-Net, Full-Deformable U-Net, and Original U-Net have all been used to assess the performance of the suggested method. The results show that, when compared to Original U-Net, the suggested technique performs better, with segmentation accuracy of 14% and classification accuracy of 99%, respectively. In a study by Raja *et al*[22] using computer-supported analysis from fundus images, a unique mechanized glaucoma identification has been carried out. A VGG-19 network design based on Support Vector Machines is used to obtain the simulation results. For glaucoma detection, the CDR threshold value of 0.41 has been employed. Fundus pictures with a CDR of 0.41 or above are glaucoma-affected, whereas those with a CDR of 0.41 or below are non-affected. The suggested glaucoma identification method utilizes widely used and attainable digital color fundus pictures. A classification precision of 94% for the collection of 175 fundus pictures has been achieved. Recently, Kumar *et al*[23] developed GlaucoNet, an ensemble DL framework (MobileNetV2+InceptionV3+ResNet50) for automated glaucoma detection from retinal fundus images. The model was trained and evaluated on two public datasets: ORIGA (650 images: 325 normal, 325 glaucoma) and ACRIMA (705 images: 309 normal, 396 glaucoma), totaling 1355 images, using 5-fold cross-validation with a 70/10/20 train/validation/test split. GlaucoNet achieved 97.0% accuracy, with 97.3% sensitivity and 96.6% specificity, supporting its potential for screening use. Gandhi *et al*[24] (2026) proposed a multi modal DL framework for glaucoma detection by integrating fundus photographs and OCT scans using a ViT architecture to capture global contextual features beyond CNNs. The model was evaluated on multiple public glaucoma datasets (reported: DRISHTI GS, ORIGA, RIGA, DRIONS DB, and GlaS) and compared against VGG16, ResNet 50, and InceptionV3. The proposed ViT framework achieved 94.5% accuracy with AUC=91.7%, outperforming the CNN baselines and supporting more interpretable clinical decision support.

Automated diagnosis of AMD utilizing fundus images[25] focuses on automated diagnosis of the afflicted macular area utilizing a hybrid feature set that combines textural, color, and structural/shape information for more precise AMD early detection. The suggested technique, in its initial step, uses the input fundus picture to identify the macular area, and then performs feature extraction based on the texture, edge, and structural characteristics of the macular region to categorize aberrant from normal macula. When used on the STARE dataset, suggested system's sensitivity, specificity, and accuracy were achieved at 97.5%, 95%, and 95.5%, respectively. Similarly, it is suggested Boukadida *et al*[26] to use an automated screening approach that can identify neovascularization from fundus photography and categories it as proliferative DR and wet AMD, or healthy. A transfer learnt model of the VGG-16 neural network is then supplied once the picture has been preprocessed for this purpose. A dataset of 395 fundus pictures of retinal images was used to assess the approach, and accuracy, sensitivity, and specificity scores of 98.30%, 98.66%, and 98.33% were attained. In another study by Xu *et al*[27], patients who visited Peking Union Medical College Hospital and had AMD or polypoidal choroidal vasculopathy (PCV) were the subject of proposed retrospective cross-sectional research. Two retinal specialists independently validated the diagnoses of each patient using the PCV and AMD diagnostic gold standards. Both spectral domain OCT pictures and color fundus images (CFI) of patients' and healthy controls' dilated eyes were collected and anonymized. Every image was pre-labeled as normal, dry, wet, or PCV AMD. As a complement to the ResNet-50 models, other ML models, such as random forest classifiers, were built for additional comparison. The bimodal CNN had the greatest performance on a test set of 143 fundus and OCT image pairings from 80 eyes (20 eyes per group), with 88.8% sensitivity and 95.6% specificity. Agarwal *et al*[71] (2026) proposed an automated dry AMD diagnosis system that avoids explicit drusen detection by focusing on the macular ROI and using handcrafted texture+color features with classical ML classifiers. The method performs macula localization, ROI selection, feature extraction, and feature selection (*t*-test and ReliefF), then trains SVM, k-nearest neighbor (KNN), Naïve Bayes classifier (NBC), and multilayer perceptron (MLP) classifiers using stratified 10-fold cross validation. It was evaluated on STARE (35 normal, 74 AMD; 109 images) and ODIR (2873 normal, 266 AMD; 3139 images) datasets. Best performance was achieved with SVM+texture features, reaching 98.89% accuracy in STARE and 95.43% in ODIR trials. Avram *et al*[72] (2026) developed a DL model to automatically identify cRORA (AMD-related atrophy) from 3D OCT volumes, reducing the burden and variability of manual grading. They compiled two independent cohorts totaling nearly 5000 OCT scans and trained a state-of-the-art volumetric model. Internal validation on the Hadassah cohort (3883 OCT volumes) achieved AUC=0.97, and external zero-shot testing on the Houston cohort (964 volumes) achieved AUC=0.88, with strong subgroup performance for non-neovascular AMD and neovascular AMD, respectively. Balaha *et al*[73] (2025) proposed a computer-aided diagnosis framework for automated eye disease classification from fundus images using ViTs and improved interpretability using SHAP explanations. The study benchmarked performance across three datasets and compared ViTs against CNN baselines. The framework reported an overall accuracy of 95%, with strong performance across multiple metrics (precision, recall, IoU, MCC), indicating potential for clinically interpretable screening support. Zedadra *et al*[74] (2025) proposed VisionTrack, a hybrid multi-modal AI system for multi-label retinal disease prediction (including AMD, DR, DME, macular hole, and drusen). The framework combines CNN-based image feature extraction with a graph neural network (GNN) for clinical risk-factor relationships and a large language model (LLM) for medical report text, and was evaluated on RFMiD (3200 fundus images) and RetinalOCT (24 000 OCT images). It achieved strong performance, reaching 0.989 accuracy on RFMiD and 0.98 accuracy on RetinalOCT, demonstrating robust generalization across modalities.

In a study on ROP by Raja Sankari *et al*[28], a hybrid DL network for ROP prediction that can used to baby mass screening was proposed. A total of 800 newborn fundus photos are used to evaluate the hybrid network after 3200 images were used to train it. The retinal vessels are separated from the fundus images using modified MultiResUNet and a matching filter with a first-order Gaussian derivative. Segmented image's contour features and grey level co-occurrence matrix are retrieved and chosen using an embedded feature selection technique. Permutation significance is used to assess the chosen features, and the random forest classifier is used to categorize them. The proposed hybrid approach outperforms ML classifiers and pre-trained models with decreased bias by predicting ROP with an accuracy of 94.5%, sensitivity of 94.0%, and specificity of 93.0%. Similarly, a cross-sectional investigation was carried out in a Taiwanese referral hospital[30]. Only preterm newborns without ROP, with ROP in stages 1 or 2, or both, were enrolled. A total of 11 372 retinal fundus images were collected and divided into 10 235 training images (90%) and 1137 images (10%) and 244 testing images. To categorize images according to the ROP stage, a deep CNN was used. Five-fold cross-validation was used to train the model, and the results showed an average accuracy of 99.93% during training and 92.23% during testing. Likewise, 5943 fundus photos from 9 North American institutions taken with a RetCam camera (Natus Medical, Pleasanton, CA) and 5049 images from four hospitals in Nepal taken with a 3nethra camera (Forus Health Incorporated, Bengaluru, India) make up two datasets in a 2022 study by Zhang *et al*[75]. On the North American test set, the CNN model that was trained on a merged dataset scored 98%.

In a recent study shared in 2024, modified Capsule Network with an improved ViT has been used to predict DR severity. This method demonstrated good performance on four datasets [APTOS, Messidor-2, deep diabetic retinopathy dataset (DDR), and EyePACS][35]. In the same year a modified VGGNet19 architecture achieved 98.84% accuracy and 100% sensitivity on (ORIGA/ORIGA-light). However, a significant limitation of this study is limited generalizability to the real-world clinical settings[36]. Later in a 2025 study, a generalized DL model VGG19-bn was trained on data from 20 different datasets and achieved exceptional performance with 98.84% accuracy. In preprocessing optic disc localization and augmentation was performed. The model accuracy dipped to 87.13 on external dataset (Drishti-GS), highlighting the need for improvement in generalizability of model[37]. In another study published in 2025, ResViT Fusion Net was proposed for DR grading. The model incorporated XAI techniques, LIME, and Grad-CAM to enhance interpret ability. Model showed excellent results with 93.01% accuracy on the APTOS 2019 dataset[38].

The progression of ophthalmic computer vision has catalyzed a shift toward automated ROP detection using DL frameworks on neonatal retinal imagery. This literature domain explores a variety of architectures targeting the precise classification and staging of infant retinal vascular abnormalities from wide-field fundus digital screening records. Current methodologies integrate hybrid convolution streams, lightweight edge-computing layers, and attention blocks to identify microvascular tortuosity and peripheral demarcation ridges[29,31,34]. A recent clinical study by Zhang *et al*[75] evaluated DL systems designed to process extensive real-world clinical records from the G-ROP Registry (7483 images) by combining image markers with text metrics, yielding an automated screening sensitivity of 96.0%. To achieve multi-class classification stability, Akbari *et al*[76] designed a hybrid DL network termed ROPDeepX that aggregates ResNet50 and EfficientNet-B4 features across 2430 images to produce a multi-class staging accuracy of 96.9% on the FARFUM dataset and a 99.70% accuracy on the cross-domain RIDIRP dataset. Parallel to server-bound systems, Mowla *et al*[31] focused on edge-computing constraints to build LightEyeNet, a lightweight mobile framework combining a DenseNet121 feature backbone with attention blocks evaluated across 1840 eHealth samples to yield a classification testing accuracy of 96.3%. Vahidmoghadam *et al*[34] (2026) developed an automated DL system for ROP screening, targeting both Plus disease detection (Plus *vs* Normal) and ROP stage classification (Stage 0–3) from retinal fundus images. Using the publicly available RIDIRP dataset from 188 infants, the study used 6004 images for Plus detection (5375 Normal; 629 Plus) and 880 images for stage classification (45 Stage 0; 252 Stage 1; 458 Stage 2; 125 Stage 3). The best-performing backbone (EfficientNetB4 with vessel-based inputs) achieved 99.6% accuracy for Plus detection and 98% accuracy for stage classification on the validation set.

The results presented in Table 1 affirm the success of AI systems in achieving remarkable performance. These results underscore the potential and effectiveness of AI in various applications and domains. The high accuracy achieved by these AI systems indicates their capability to make accurate predictions and classifications, thus highlighting their relevance and value in real-world scenarios.

Status of Integration of AI Systems into Clinical Practice

However, it has been empirically acknowledged and proved that AI systems, when properly trained and validated, have the potential to achieve detection limits comparable to or even surpassing those of human experts[25-26]. Despite the promising performance of AI algorithms in automated disease detection, several barriers exist that hinder their translation into clinical practice[77]. This section aims to explore and analyze the barriers that impede the translation of CAD systems into clinical practice. By examining existing research and scholarly articles, this review seeks to provide insights into the challenges and complexities surrounding CAD implementation, shedding light on key areas such as validation and generalizability, regulatory considerations, integration with clinical workflows, interpretation of CAD output, user acceptance and trust, as well as cost and resource considerations. Understanding these barriers is vital for devising strategies and guidelines to facilitate the effective and efficient utilization of CAD systems in routine clinical practice. Abramoff *et al*[78] highlighted some significant scientific and nonscientific challenges to the use of automated DR detection in clinical practice in preliminary research published in 2010. Scientific concerns including disease detection limitations as comparison to human specialists and nonscientific barriers like ethical, legal, and political concerns. It has been determined that to better manage translation into clinical practice, these obstacles must be measured and addressed. In recent research, Burlina *et al*[79] highlighted a potential cause for the performance degradation of AI systems in real-time clinical settings as the absence of publicly accessible huge datasets containing a varied range of samples. Similarly, other research has been observed that performance of AI systems is negatively impacted by data from a single-center database that is skewed due to factors like patient race and imaging technology[80]. Likewise, one of the main reasons why ophthalmologists do not depend on such tools is that the lack of clinically relevant information that physicians believe is essential for evaluating efficacy is commonly absent from a computational standpoint on AI and especially in DL systems[81].

Successful clinical adoption of AI-based ocular diagnostic systems requires close collaboration between clinicians and researchers throughout the development cycle. Clinicians can help define the intended use (screening *vs* triage *vs* diagnosis), identify clinically meaningful outputs and referral thresholds, and ensure that model evaluation metrics reflect real clinical needs. Researchers, in collaboration with clinicians, can establish robust labeling and adjudication protocols (*e.g.*, multi-grader consensus), design appropriate validation strategies including external and multi-center testing, and conduct prospective clinical studies to assess real-world performance. Finally, clinician-in-the-loop feedback during deployment is important for identifying failure cases, improving usability, and supporting continuous monitoring and safe model updates.

AI offers several benefits in ocular disease diagnosis, particularly for large-scale screening programs, including high diagnostic accuracy, consistent grading, reduced workload for specialists, and earlier detection of diseases such as DR and glaucoma as shown in Table 1. However, it is important to differentiate between accuracy reported in retrospective research settings and reliability in real-world clinical practice. Many models achieve excellent performance on curated datasets, but their performance may degrade in routine settings due to variations in image quality, different cameras/devices, differences in patient populations, and workflow-related factors such as ungradable images and delays in referral pathways. Therefore, high accuracy alone does not guarantee clinical reliability unless models undergo robust external validation on independent datasets, are monitored after deployment to detect performance drift, and are integrated into clinical workflows with clear referral thresholds and appropriate clinician oversight.

Smartphone Fundoscopy as a User-Friendly and Cost-Effective Alternative to Conventional Digital Cameras for Capturing Fundus Images This section explores the potential of smartphone fundoscopy, highlighting its ability to democratize retinal health assessment, facilitate telemedicine, empower patients, and contribute to early detection and intervention of retinal diseases. A study to validate the quality and utility of fundus images captures using smartphone camera and digital fundus camera has been conducted[30]. A total of 2152 images were obtained from both methods. Both image sets were independently graded at Moorfields Eye Hospital Reading Centre. It was concluded that fundus images using smartphone can be effectively used for retinal diseases analysis. A comparable validation of fundus photos taken with a mobile camera and a fundus camera was done in different research to identify DR on both types of images[75]. It was determined that retinal photography utilizing a mobile camera has significant agreement with conventional retinal photography and is highly successful for the screening and diagnosis of DR. In similar study, Hacisoftaoglu *et al*[41] looked into the data collection methods and publicly accessible smartphone-based retinal imaging systems, such as iExaminer, D-Eye, Peek Retina, and iNview, to see if they were appropriate for research on smartphone-based retinal imaging. According to the findings, when compared to the retinal imaging systems iExaminer, D-Eye, and Peek Retina, the iNview system offers the biggest field of view and the best picture quality. The field of vision of D-Eye and Peek Retina is about the same size and is half that of iNview. The shortest field of view is found in iExaminer. For iExaminer, D-Eye, and Peek Retina, the optimum field of view is attained when photos are taken at the shortest distance (22 mm). However, at a distance of 65 mm, iNview takes the highest-quality photos. Similarly, non-contact smartphone fundus images and conventional CFI were evaluated by Wintergerst *et al*[82] for their suitability for ROP screening and documentation. smartphone fundus images and CFI were taken of 26 eyes. Likewise, a recent study strongly suggests use of smartphone-based imaging for early diagnosis of retinal diseases. It was empirically investigated and determined that smartphone fundus imaging is a non-contact, affordable substitute for CFI for ROP screening and documentation that has the potential to significantly enhance ROP care in settings with limited resources[37]. According to the literature and empirical investigations, smartphone fundoscopy offers a convenient and affordable alternative to traditional digital cameras for taking fundus photographs. Recent work has also proposed low-cost AI-enabled screening systems targeting resource-limited settings. For example, Vohra *et al*[63] presented a cost-effective fundus imaging setup combined with CNN-based multi-disease detection (DR/AMD/glaucoma) and a cloud-based reporting workflow, reporting high diagnostic performance. This supports the growing interest in practical, low-cost deployment models for underserved regions, which aligns with our Barrier 1 discussion.

AI-Based Retinal Screening Algorithms using Smartphone Fundoscopy This section aims to evaluate the performance and accuracy of AI-based retinal screening algorithms using smartphone fundoscopy. By examining the existing literature and studies, this section will assess the effectiveness of these algorithms in detecting various retinal diseases.

Mrad *et al*[39] used fundus images captured using smartphone associated to an optical lens. Segmented vessel tree is partitioned into quadrant based on the inferior–superior–nasal–temporal (ISNT) rule. The centroid of each quadrant is measured based on coordination of white pixels. Finally, the android software development kit (SDK) is used to deploy this MATLAB code on smartphone. The proposed system achieved 99% accuracy, 96.77% specificity, and 100% sensitivity on publicly available DRISHTI-DB dataset. It also showed outstanding results with 100% accuracy on custom dataset of 24 images taken from smart phone.

Rajalakshmi *et al*[40] conducted a study to evaluate the use of automated AI-based software for the identification of DR and sight-threatening DR (STDR) using fundus photography captured with a smartphone-based device and to compare the results to ophthalmologist grading. Using digital and smartphone fundus cameras, fundus photos of 301 patients were taken. The ophthalmologists graded the DR using the international clinical DR classification system. Proliferative DR (PDR), DME, or severe non-proliferative DR were used to define STDR. A proven AI DR screening tool (EyeArtTM: developed by Eyenuk Inc.) that can distinguish between DR, referable DR (moderate non-proliferative DR or worse and/or DME), and STDR was used to assess the retinal images. The AI software showed 95.8% sensitivity and 80.2% specificity for detecting any DR and 99.1% sensitivity and 80.4% specificity in detecting STDR. Results demonstrate that AI analysis of smartphone retinal imaging has a very high sensitivity for identifying DR and STDR, it can be used as a first step in mass retinal screening of diabetics.

Using a DL strategy and the ResNet50 network, Hacisoftaoglu *et al*[41] suggested an automated DR detection model for smartphone-based retinal images. The transfer learning methodology was originally used to the well-known AlexNet, GoogLeNet, and ResNet50 architectures. Second, to test the effects of employing images from a single, cross, and multiple datasets, these frameworks were retrained using 45 000 retina images from several datasets. Third, to investigate the DR detection precision of smartphone-based retinal imaging systems, the suggested ResNet50 model is applied to smartphone-based synthetic pictures. Since no smartphone-based retinal imaging dataset is publicly available, synthetic retina images were created by modelling the field of vision for several smartphone-based devices using the original retina images from publicly available datasets. The suggested technique has a high classification accuracy of 98.6%, a 98.2% sensitivity, and a 99.1% specificity. Similarly, in a remote field setting in Mexico, a study evaluated two AI algorithms (Medios offline and EyeArt online) trained on standard fundus images for the DR screening using fundus images taken with a smart phone. Both of the AI systems demonstrated high sensitivity (94%–99%) and specificity (86%–94%) that were on a par with clinical standards, despite issues like operator errors[42].

It is quite evident from Table 2 that despite the limited advancements in the field of AI-based retinal screening algorithms using smartphone fundoscopy, the performance of existing solutions in this domain shows promising results. The combination of AI algorithms and smartphone technology has the potential to transform retinal disease screening and improve access to care. Literature has also revealed a significant scarcity of publicly accessible, standardized dataset for smartphone fundus images, as used by Mrad *et al*[39] and Rajalakshmi *et al*[40], whereby a very small dataset of smartphone fundus images has been used. However, a study by Hacisoftaoglu *et al*[41] used synthetic smartphone images which have been generated through digital fundus images.

Algorithm Optimization Opportunities for Running Deep Learning and AI Models on Edge Devices The main objective of this section is to explore strategies and techniques for optimizing algorithms specifically tailored to enable efficient execution of DL and AI models on edge devices.

In the 2012 ImageNet Challenge, Krizhevsky *et al*[83] produced ground-breaking results with a network with 60 million parameters, 5 convolutional layers, and 3 fully connected layers. On the ImagetNet data set, training the whole model with an NVIDIA K40 computer typically takes 2 to 3d. Non-saturating neurons and a very effective GPU implementation of convolutional networks were employed to speed up training. They also used a novel regularization technique to greatly decrease overfitting in the globally linked layers.

Model compression is a technique used in algorithm optimization to reduce the size and computational complexity of DL or AI models without significantly sacrificing their performance. In a recent study, for deep CNN-based speech enhancement, which combines three separate techniques sparse regularization, iterative pruning, and clustering-based quantization, Tan and Wang[43] presented two compression methods. These methods were thoroughly investigated, and the suggested compression pipelines were assessed. Experimental findings show that the suggested greatly shrinks the diameters of four distinct models without significantly compromising their enhancing efficacy. Similarly, NVIDIA Jetson Tx2 is a typical embedded DL architecture, and Qin *et al*[44] establish a quantitative way to characterize model compression methods on it. Numerous tests were carried out by taking into account 11 significant neural network designs from the image classification and natural language processing fields. The effectiveness of two widely used compression methods, data quantization and pruning, on various network topologies as well as the effects of compression on model storage size, inference time, energy consumption, and performance indicators were empirically demonstrated.

In order to accomplish in-network compression in the early network layers, it was suggested in by Matsubara *et al*[45] that the structure and training procedure of CNN models for complicated image classification tasks be modified. The training procedure is based on knowledge distillation, a method that has historically been used to create tiny student models that imitate the results of larger instructor models. This concept is used in this instance to achieve extreme compression while maintaining accuracy. Results show that the suggested technique works well for cutting-edge models developed across complicated datasets, and it can expand the parameter range where edge computing is a practical and desirable choice. Similarly in a recent study, PSO, genetic algorithms (GA), and simulated annealing algorithms (SAA) have been analyzed in a study by You and Tang[46] for algorithm optimization. These algorithms model the task offloading problem in the industrial internet of things (IoT) environment as a multi-user and multi-mobile edge computing (MEC) problem. A penalty function was included to balance the energy usage and delay to get rid of the job processing queuing delay. A comparison of PSO-based offloading approach with the GA-based offloading strategy and the SAA-based dumping strategy was made. According to the study of experimental findings, the strategy of PSO outperforms the strategies of GA and SAA as the number of tasks and equipment increases. Similarly, using the Levy flight approach, a discrete variant of the butterfly optimization algorithm that accelerates convergence and guards against local optima issues has been presented by Hosseinzadeh *et al*[47]. To determine the job execution order in the scientific processes, a task prioritization approach was also used. Then, in MEC contexts, dynamic voltage and frequency scaling-based data-intensive process scheduling and data placement was employed. Extensive simulations are performed on several well-known scientific processes of varying sizes to assess the performance of the proposed scheduling method. The results of the experiment show that the proposed approach can perform better than existing algorithms in terms of energy use, data access overheads, and other factors.

Parallelization and concurrency are powerful techniques in algorithm optimization that can significantly improve performance by leveraging multiple processing units or threads to execute tasks concurrently. A formal model, a VHSIC hardware description language (VHDL) model, techniques, and software for automated parallelization of algorithms run on an asynchronous network were provided by Prihozhy *et al*[48]. A group of concurrent operation and variable pair definitions determine the amount of concurrency of the net algorithm. The methods and tools are aimed towards asynchronous high- and system-level information processing synthesis and optimization in computer networks. Likewise, edge computers' central processing is getting more significant. Yoo[49] centered his attention on ARM's power-performance scaling processor design to suit a variety of workload characteristics. The design combines energy-efficient and high-performance cores to flexibly scale power and performance as needed. The two power-performance scaling techniques shown here are based on achievable throughput and typical power usage.

On-device training refers to the process of training ML models directly on edge devices, such as smartphones. The approach suggested by Jang *et al*[50] enables knowledge transfer and policy model compression in a single training procedure on edge devices while considering their constrained resource budgets. The distinctive aspect of the suggested methodology is that it uses a knowledge distillation method to manage edge devices in integrated edge cloud settings. Analysis was done on the performance of the suggested solution when it was implemented on a commercial embedded system-on-module with constrained hardware resources. The experimental results demonstrate that: 1) the proposed method for edge policy training achieves near-cloud performance in terms of average rewards, even though the size of the edge policy network is significantly smaller than that of the cloud policy network; 2) the training time for edge policy training is significantly shorter than that for cloud policy training. Likewise, a unique backpropagation architecture and a low-power, low-resource algorithm is proposed by Kaneko *et al*[51]. Both the forward pass and the reverse pass are calculated using fixed-point multiplications. The reverse pass is calculated using terrorized gradient. The reverse pass is calculated using ternarized gradient. It has been demonstrated empirically that the suggested technique uses two orders of magnitude less power than 16-bit quantized backpropagation on the identical objective, which is a two-class classification problem using a separate dataset.

The status of algorithm optimization opportunities for implementing AI models directly on edge devices has shown significant advances in recent years[84]. The authors discuss technical challenges like resource constraints and offer optimization techniques like hardware acceleration, model

Table 6 A summary of barriers, solutions, and future directions

Barriers	Key findings	Future directions
B1: Clinical integration&cost	High complexity, GPU dependence, and workflow mismatches limit adoption	Develop smartphone/edge-AI solutions (*e.g.*, fundoscopy apps)–Improve EHR/PACS interoperability–Advocate regulatory sandboxes for faster approvals
B2: Lack of transparency	"Black-box" AI reduces clinician trust and raises ethical concerns	Adopt XAI techniques (LIME, SHAP, Grad-CAM)–Implement model cards for bias/performance disclosure–Hybrid clinician-AI workflows for oversight
B3: Poor generalization	Biases in training data worsen outcomes for underrepresented groups	Multicentric datasets (*e.g.*, federated learning)–Synthetic data augmentation (GANs for rare cases)–Subgroup-specific validation (age, ethnicity, device type)
Cross-cutting themes	AI performs well in research (>90% accuracy) but falters in real-world use	Real-world validation (longitudinal studies)–Global equity initiatives (LMIC-focused deployments)–Dynamic model updates *via* clinician feedback

AI: Artificial intelligence; EHR: Electronic health record; GAN: Generative adversarial network; GPU: Graphics processing unit; Grad-CAM: Gradient-weighted class activation mapping; LIME: Local interpretable model-agnostic explanations; LMIC: Low- and middle-income country; PACS: Picture archiving and communication system; SHAP: SHapley Additive exPlanations; XAI: Explainable artificial intelligence.

compression, and quantization. Algorithm optimization is considered essential for enabling efficient execution of AI models on edge devices, which typically have limited computational resources. To bridge the clinical translation gap, recent research highlights the necessity of hardware-aware, feature-driven attention mechanisms and lightweight, multi-stage Transformer architectures for efficient, interpretable on-device processing[52-53]. To address institutional domain shifts and data-sharing constraints, recent 2026 frameworks employ decentralized optimization and contrastive pretraining architectures to maximize cross-site model generalization. Decentralized optimization frameworks, such as parallel networks refined by nature-inspired metaheuristic convergence loops, significantly enhance multi-center model stability and training speeds without requiring raw repository centralization[54]. Simultaneously, federated contrastive pretraining enforces localized weight alignment during collaborative training phases, pushing deep architectures to capture invariant structural anomalies instead of site-specific image noise[55]. This combined dual paradigm provides a mathematically verified pathway for ophthalmic software to maintain high diagnostic precision when deployed across entirely unseen, foreign diagnostic registries.

Algorithm optimization techniques for edge AI deployment are summarized in Table 3. The status of algorithm optimization opportunities for running DL and AI models on edge devices has shown significant advancements in recent years. As edge devices, such as smartphones, IoT devices, and embedded systems, become more powerful, there is an increasing demand for deploying sophisticated AI models directly on these devices to enable real-time and privacy-preserving applications[16]. Algorithm optimization plays a crucial role in achieving efficient execution on edge devices with limited computational resources.

CONCLUSION

AI has the potential to revolutionize ocular diagnostics, but its success hinges on addressing technical, clinical, and ethical challenges. By prioritizing generalization, transparency, and seamless integration, stakeholders can unlock AI's full potential to reduce preventable vision loss worldwide. Future research must bridge the "last-mile gap" between bench and bedside, ensuring AI tools are not only accurate but also equitable, trusted, and actionable in real-world practice. Future research should focus on prospective multi-center validation, stronger fairness and generalizability testing across populations and devices, and practical deployment through low-cost smartphone/edge-AI systems for resource-limited settings. Studies should also include post-deployment monitoring and measure real clinical impact (*e.g.*, referral uptake and time-to-treatment), not accuracy alone. Table 6 summarizes the key findings of the study outlining barriers, solutions, and future directions.

ACKNOWLEDGEMENTS

Authors' Contributions: Shehryar T conceived the study, supervised the project, and critically revised the manuscript; Shehryar T and Fazal ZZ performed the literature review, data extraction, data analysis, figure preparation, and drafted the manuscript; Salam AA contributed to data extraction and manuscript revision; Varada S contributed to data analysis and manuscript revision; Jabran F assisted with data interpretation and manuscript revision; Akram MU supervised the artificial intelligence methodology and critically revised the manuscript. All authors reviewed and approved the final manuscript.

AI-Generated Content Disclosure: No artificial intelligence tools were used in the preparation of this manuscript.

Conflicts of Interest: Shehryar T, None; **Fazal ZZ,** None; **Salam AA,** None; **Varada S,** None; **Jabran F,** None; **Akram MU,** None.

REFERENCES

1 World Health Organization. World report on vision. Geneva, Switzerland: World Health Organization; 2019. https://www.who.int/publications/i/item/world-report-on-vision

2 Hagiwara Y, Ciora OA, Monnet M, *et al*. AI-driven approaches for glaucoma detection—a comprehensive review. 2024:2410.15947. https://arxiv.org/abs/2410.15947

3 Ennab M, Mcheick H. Enhancing interpretability and accuracy of AI models in healthcare: a comprehensive review on challenges and future

directions. *Front Robot AI* 2024;11:1444763.

4 Amann J, Blasimme A, Vayena E, *et al*. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak* 2020;20(1):310.

5 Wang F, Beecy A. Implementing AI models in clinical workflows: a roadmap. *BMJ Evid Based Med* 2025;30(5):285-287.

6 Adler-Milstein J, Aggarwal N, Ahmed M, *et al*. Meeting the Moment: addressing barriers and facilitating clinical adoption of artificial intelligence in medical diagnosis. NAM Perspectives. Washington, DC: National Academy of Medicine; 2022.

7 Olawade DB, David-Olawade AC, Wada OZ, *et al*. Artificial intelligence in healthcare delivery: Prospects and pitfalls. *J Med Surg Public Heal* 2024;3:100108.

8 Rajesh AE, Lee AY, Specialist R. AI for DR screening: Where are we in 2025? 2025. http://www.retina-specialist.com/article/ai-for-dr-screening-where-are-we-in-2025

9 Tertel Z. AI-based blink rate model may help predict and detect dry eye disease early. *Ophthalmology Advisor* 2025. https://www.ophthalmologyadvisor.com/home/topics/dry eye/ai based blink rate model may help predict and detect dry eye disease early/

10 Skevas C, Weindler H, Levering M, *et al*. Simultaneous screening and classification of diabetic retinopathy and age-related macular degeneration based on fundus photos-a prospective analysis of the RetCAD system. *Int J Ophthalmol* 2022;15(12):1985-1993.

11 Ashikur M, Arifur M, Ahmed J. Automated detection of diabetic retinopathy using deep residual learning. *Int J Comput Appl* 2020;177(42):25-32.

12 Sudha S, Srinivasan A, Devi TG. Cross-validation convolution neural network-based algorithm for automated detection of diabetic retinopathy. *Comput Syst Sci Eng* 2023;45(2):1985-2000.

13 Madi MMM, Farr PC, Bester D. Diabetic retinopathy detection: AI models and approaches. *J Ophthalmol* 2026;2026:8857887.

14 Kour M, Saini K. A deep learning framework of multi-class diabetic retinopathy detection with an explainable attention. *2026 9th International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech)* Kolkata, India. IEEE, 2026;1-6.

15 Al-Shalabi L, Al-Shalabi R. A dependable approach for the early detection of diabetic retinopathy using hybrid machine learning models. *J Electr Comput Eng* 2026;2026:9298734.

16 Cao K, Xu J, Zhao WQ. Artificial intelligence on diabetic retinopathy diagnosis: an automatic classification method based on grey level co-occurrence matrix and naive Bayesian model. *Int J Ophthalmol* 2019;12(7):1158-1162.

17 Sundaram S, Selvamani M, Raju SK, *et al*. Diabetic retinopathy and diabetic macular edema detection using ensemble based convolutional neural networks. *Diagnostics* (*Basel*) 2023;13(5):1001.

18 Velcani F. AI tool finds retinal biomarkers that predict cognitive impairment risk in diabetes. *Ophthalmology Advisor* 2026. Accessed on: June 1, 2026. https://www.ophthalmologyadvisor.com/reports/artificial-intelligence-retinal-biomarkers-predict-cognitive-impairment-risk-in-diabetes/

19 Singh LK, Khanna M, Thawkar S, *et al*. Collaboration of features optimization techniques for the effective diagnosis of glaucoma in retinal fundus images. *Adv Eng Softw* 2022;173:103283.

20 Shoukat A, Akbar S, Al E Hassan S, *et al*. An automated deep learning approach to diagnose glaucoma using retinal fundus images. *2021 International Conference on Frontiers of Information Technology (FIT)*. Islamabad, Pakistan. IEEE, 2021;120-125.

21 Zedan MJM, Zulkifley MA, Ibrahim AA, *et al*. Automated glaucoma screening and diagnosis based on retinal fundus images using deep learning approaches: a comprehensive review. *Diagnostics* (*Basel*) 2023;13(13):2180.

22 Raja J, Shanmugam P, Pitchai R. An automated early detection of glaucoma using support vector machine based visual geometry group 19 (VGG-19) convolutional NeuralNetwork. *Wirel Pers Commun* 2021;118(1):523-534.

23 Kumar M, Dembla D, Goyal V. Artificial intelligence-driven glaucoma screening in ophthalmology: The GlaucoNet deep learning framework. *Afr Vis Eye Heal* 2026;85:a1083.

24 Gandhi VC, Gandhi PP, Abdul Raheem AK, *et al*. Advancing glaucoma diagnosis: multi-modal deep learning with vision transformer architectures. *Intell Based Med* 2026;13:100355.

25 Khalid S, Akram MU, Shehryar T, *et al*. Automated diagnosis system for age-related macular degeneration using hybrid features set from fundus images. *Int J Imaging Syst Tech* 2021;31(1):236-252.

26 Boukadida R, Elloumi Y, Kachouri R, *et al*. Automated diagnosis of retinal neovascularization pathologies from color retinal fundus images. *Advances in Computer Graphics*. Cham: Springer Nature Switzerland; 2022:451-462.

27 Xu ZY, Wang WS, Yang JY, *et al*. Automated diagnoses of age-related macular degeneration and polypoidal choroidal vasculopathy using bi-modal deep convolutional neural networks. *Br J Ophthalmol* 2021;105(4):561-566.

28 Raja Sankari V, Snekhalatha U, Chandrasekaran A, *et al*. Automated diagnosis of Retinopathy of prematurity from retinal images of preterm infants using hybrid deep learning techniques. *Biomed Signal Process Control* 2023;85:104883.

29 Tong Y, Lu W, Deng QQ, *et al*. Automated identification of retinopathy of prematurity by image-based deep learning. *Eye Vis (Lond)* 2020;7:40.

30 Chen JS, Coyner AS, Ostmo S, *et al*. Deep learning for the diagnosis of stage in retinopathy of prematurity: accuracy and generalizability across populations and cameras. *Ophthalmol Retina* 2021;5(10):1027-1035.

31 Mowla N, Mowla MN, Rabie K, *et al*. A lightweight deep learning model for retinopathy of prematurity classification in eHealth applications. *2025 International Wireless Communications and Mobile Computing (IWCMC)*. Abu Dhabi, United Arab Emirates, 2025;227-232.

32 Zhao X, Wu Z, Wu S, *et al*. Multimodal artificial intelligence in retinopathy of prematurity: A comprehensive narrative review. *Surv Ophthalmol* 2026;S0039-6257(26)00063-9.

33 Mohiy E, AbdulWakel HI, Khairy M, *et al*. An efficient deep learning model for reliable detection and classification of retinopathy of prematurity. *Discover Artif Intell* 2026;6(1):305.

34 Vahidmoghadam M, Ghorbani P, Ahmadi MJ, *et al*. Automated diagnosis of plus form and early stages of ROP using deep learning models. *Sci Rep* 2026;16:7234.

35 Oulhadj M, Riffi J, Khodriss C, *et al*. Diabetic retinopathy prediction based on vision transformer and modified capsule network. *Comput Biol Med* 2024;175:108523.

36 Tampa H, Mekongo M, Tiedeu A. Deep learning-based algorithm for automated detection of glaucoma on eye fundus images. *Multimed Tools Appl* 2025;84(20):22809-22826.

37 Chaurasia AK, Liu GS, Greatbatch CJ, *et al*. A generalised computer vision model for improved glaucoma screening using fundus images. *Eye (Lond)* 2025;39(1):109-117.

38 Ikram A, Imran A. ResViT FusionNet Model: an explainable AI-driven approach for automated grading of diabetic retinopathy in retinal images. *Comput Biol Med* 2025;186:109656.

39 Mrad Y, Elloumi Y, Akil M, *et al*. A fast and accurate method for glaucoma screening from smartphone-captured fundus images. *Innovation and Research in BioMedical Engineering* 2022;43(4):279-289.

40 Rajalakshmi R, Subashini R, Anjana RM, *et al*. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. *Eye* (*Lond*) 2018;32(6):1138-1144.

41 Hacisoftaoglu RE, Karakaya M, Sallam AB. Deep learning frameworks for diabetic retinopathy detection with smartphone-based retinal imaging systems. *Pattern Recognit Lett* 2020;135:409-417.

42 Wroblewski JJ, Sanchez-Buenfil E, Inciarte M, *et al*. Diabetic retinopathy screening using smartphone-based fundus photography and deep-learning artificial intelligence in the yucatan peninsula: a field study. *J Diabetes Sci Technol* 2025;19(2):370-376.

43 Tan K, Wang DL. Towards model compression for deep learning based speech enhancement. *IEEE/ACM Trans Audio Speech Lang Process* 2021;29:1785-1794.

44 Qin Q, Ren J, Yu JL, *et al*. To compress, or not to compress: characterizing deep learning model compression for embedded inference. 2018;arXiv:1810.08899.

45 Matsubara Y, Callegaro D, Baidya S, *et al*. Head network distillation: splitting distilled deep neural networks for resource-constrained edge computing systems. *IEEE Access* 2020;8:212177-212193.

46 You Q, Tang B. Efficient task offloading using particle swarm optimization algorithm in edge computing for industrial Internet of Things. *J Cloud Comput* 2021;10(1):41.

47 Hosseinzadeh M, Masdari M, Rahmani AM, *et al*. Correction to: improved butterfly optimization algorithm for data placement and scheduling in edge computing environments. *J Grid Comput* 2021;19(3):27.

48 Prihozhy A, Merdjani R, Iskandar F. Automatic parallelization of net algorithms. *Proceedings International Conference on Parallel Computing in Electrical Engineering. PARELEC 2000* Trois-Rivieres, QC, Canada. IEEE, 2000:24-28.

49 Yoo S. Leveraging multicores for mobile edge computing. *2018 International Conference on Information Networking (ICOIN)* Chiang Mai, Thailand. IEEE, 2018:869-874.

50 Jang I, Kim H, Lee D, *et al*. Knowledge transfer for on-device deep reinforcement learning in resource constrained edge computing systems. *IEEE Access* 2020;8:146588-146597.

51 Kaneko T, Orimo K, Hida I, *et al*. A study on a low power optimization algorithm for an edge-AI device. *Nonlinear Theory and Its Applications, IEICE* 2019;10(4):373-389.

52 Gurnani B, Kaur K. Artificial intelligence in ophthalmology: from innovation to clinical integration. *Front Ophthalmol* 2026;6:1839194.

53 Altayeb K, Ansari S, Bonny T, *et al*. A novel explainable AI framework for multi-disease ocular classification and diabetic retinopathy severity grading. *Neural Comput Appl* 2026;38(8):251.

54 Baddur K, Sangaralingam P. Federated learning for diabetic retinopathy detection using parallel convolutional LeNet. *Knowl Based Syst* 2026;342:115908.

55 Cacciatore A, Di Cosmo M, Frontoni E, *et al*. Federated learning towards the Unknown: a deep dive into diabetic retinopathy prediction from Real-World ehr structured data on Unseen diabetic centers. *Machine Learning and Knowledge Discovery in Databases. Applied Data Science Track*. Cham: Springer Nature Switzerland; 2026:338-355.

56 Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": explaining the predictions of any classifier. 2016;arXiv:1602.04938.

57 Wachter S, Mittelstadt B, Floridi L. Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *Int Data Priv Law* 2017;7(2):76-99.

58 Adeniran AA, Onebunne AP, William P. Explainable AI (XAI) in healthcare: enhancing trust and transparency in critical decision-making. *World J Adv Res Rev* 2024;23(3):2447-2658.

59 Burrell J. How the machine 'thinks': understanding opacity in machine learning algorithms. *Big Data Soc* 2016;3(1):2053951715622512.

60 Lundberg S, Lee SI. A unified approach to interpreting model predictions. 2017:1705.07874. https://arxiv.org/abs/1705.07874

61 Selvaraju RR, Cogswell M, Das A, *et al*. Grad-CAM: visual explanations from deep networks *via* gradient-based localization. *Int J Comput Vis* 2020;128(2):336-359.

62 Mitchell M, Wu S, Zaldivar A, *et al*. Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency* 2019.

63 Vohra H, Hasan MK, Norul Huda Sheikh Abdullah S, *et al*. A low-cost AI-powered system for early detection of diabetic retinopathy and ocular diseases in resource limited settings. *IEEE Access* 2025;13:97322-97336.

64 Dutt S, Sivaraman A, Savoy F, *et al*. Insights into the growing popularity of artificial intelligence in ophthalmology. *Indian J Ophthalmol* 2020;68(7):1339-1346.

65 Luo Y, Tian Y, Shi M, *et al*. Harvard glaucoma fairness: a retinal nerve disease dataset for fairness learning and fair identity normalization. *IEEE Trans Med Imaging* 2024;43(7):2623-2633.

66 Liu BY, Chakor H, Kobbi R, *et al*. A domain adaptation method for deep learning based automatic diabetic retinopathy grading. *Invest Ophthalmol Vis Sci* 2023;64(8):263.

67 Machado J, Marta A, Mestre P, *et al*. Data augmentation with generative methods for inherited retinal diseases: a systematic review. *Appl Sci* 2025;15(6):3084.

68 Herlihy C, Truong K, Chouldechova A, *et al*. A structured regression approach for evaluating model performance across intersectional subgroups. *arXiv* 2024;2401.14893.

69 Hallaj S, Chuter BG, Lieu AC, *et al*. Federated learning in glaucoma: a comprehensive review and future perspectives. *Ophthalmol Glaucoma* 2025;8(1):92-105.

70 Edmonds J, Chan RVP, Yi D. Expert in the loop training to optimize clinician time. *Invest Ophthalmol Vis Sci*. 2022;63(7):193-F0040.

71 Agarwal D, Bhargava A, Alsharif MH, *et al*. Automatic diagnosis of age-related macular degeneration using machine learning and image processing techniques. *Sci Rep* 2026;16(1):5037.

72 Avram O, Shwartz Y, Green A, *et al*. A deep learning model for automated identification of age-related macular degeneration atrophy. *Graefes Arch Clin Exp Ophthalmol* 2026. Epub ahead of print. PMID:42105087.

73 Balaha HM, Hassan AE, Ahmed RA, *et al*. Advancing eye disease detection: a comprehensive study on computer-aided diagnosis with vision transformers and SHAP explainability techniques. *Biocybern Biomed Eng* 2025;45(1):23-33.

74 Zedadra A, Salah-Salah MY, Zedadra O, *et al*. Multi-modal AI for multi-label retinal disease prediction using OCT and fundus images: a hybrid approach. *Sensors* (*Basel*) 2025;25(14):4492.

75 Zhang GH, Lin JW, Wang J, *et al*. Automated multidimensional deep learning platform for referable diabetic retinopathy detection: a multicentre, retrospective study. *BMJ Open* 2022;12(7):e060155.

76 Akbari M, Pourreza HR, Khalili Pour E, *et al*. FARFUM-RoP, A dataset for computer-aided detection of retinopathy of prematurity. *Sci Data* 2024;11:1176.

77 Jin K, Ye J. Artificial intelligence and deep learning in ophthalmology: Current status and future perspectives. *Adv Ophthalmol Pract Res* 2022;2(3):100078.

78 Abramoff MD, Niemeijer M, Russell SR. Automated detection of diabetic retinopathy: barriers to translation into clinical practice. *Expert Rev Med Devices* 2010;7(2):287-296.

79 Burlina P, Paul W, Mathew P, *et al*. Low-shot deep learning of diabetic retinopathy with potential applications to address artificial intelligence bias in retinal diagnostics and rare ophthalmic diseases. *JAMA Ophthalmol* 2020;138(10):1070-1077.

80 Tan TN, Anees A, Chen C, *et al*. Retinal photograph-based deep learning algorithms for myopia and a blockchain platform to facilitate artificial intelligence medical research: a retrospective multicohort study. *Lancet Digit Heal* 2021;3(5):e317-e329.

81 CONSORT-AI and SPIRIT-AI Steering Group. Reporting guidelines for clinical trials evaluating artificial intelligence interventions are needed. *Nat Med* 2019;25(10):1467-1468.

82 Wintergerst MWM, Petrak M, Li JQ, *et al*. Non-contact smartphone-based fundus imaging compared to conventional fundus imaging: a low-cost alternative for retinopathy of prematurity screening and documentation. *Sci Rep* 2019;9(1):19711.

83 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Comm ACM* 2017;60(6):84-90.

84 Wang XB, Tang ZQ, Guo JX, *et al*. Empowering edge intelligence: a comprehensive survey on on-device AI models. *ACM Comput Surv* 2025;57(9):1-39.