© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722392

Artificial Intelligence-Assisted Interpretation of Ophthalmic Images for Early Detection of Retinal Diseases

MUHAMMAD QASIM
0009-0005-2541-8176

Abstract- ***Retinal diseases constitute major causes of avoidable visual impairment, with early manifestations that are frequently subtle, spatially distributed, and challenging to interpret consistently, particularly in busy or under-resourced clinical settings. Artificial intelligence-assisted interpretation of ophthalmic images enables earlier recognition by transforming fundus photographs, optical coherence tomography, optical coherence tomography angiography, and multimodal records into reproducible estimates of disease presence, severity, and referral urgency. This review thoroughly examines evidence published from 2020 to 2025 regarding image-based artificial intelligence for early detection of diabetic retinopathy, age-related macular degeneration, glaucoma, retinal vascular occlusion, and inherited retinal degeneration. In contrast to broad reviews of artificial intelligence in ophthalmology, this work focuses on the entire early-detection pathway, including image acquisition, quality control, lesion localization, diagnostic classification, uncertainty communication, clinical triage, and longitudinal monitoring. A structured narrative review was carried out using PubMed, Scopus, Web of Science, IEEE Xplore, and ScienceDirect, employing targeted searches that combined retinal disease terms, imaging modalities, artificial intelligence methods, screening, validation, explainability, and clinical deployment. Evidence was synthesized with respect to diagnostic performance, transportability, workflow value, and patient-safety requirements. While deep learning systems are able to achieve high discrimination in selected datasets, their real-world value depends on image quality, spectrum diversity, external validation, calibration, interoperable reporting, and clinically meaningful thresholds. Fundus photography remains highly scalable for population screening, whereas optical coherence tomography and multimodal models provide more comprehensive structural and temporal information for macular and glaucomatous disease. The review concludes that artificial intelligence should serve as a supervised interpretation layer rather than an autonomous replacement for clinical judgment. Upcoming studies must prioritize prospective multicenter evaluation, subgroup fairness, uncertainty-aware outputs, privacy-preserving learning, and outcome-based comparisons that assess earlier treatment, preserved vision, and reduced inequity rather than accuracy alone.***

Keywords: artificial intelligence; retinal imaging; early detection; fundus photography; optical coherence tomography; diabetic retinopathy; age-related macular degeneration; glaucoma; explainable AI

I. INTRODUCTION

Retinal disease provides a unique opportunity for artificial intelligence, as diagnosis is highly dependent on imaging and many clinically significant abnormalities are detectable before patients experience symptoms. Microaneurysms, small intraretinal hemorrhages, drusen, pigmentary disturbances, localized retinal nerve fiber layer loss, vascular nonperfusion, and subtle photoreceptor disruptions may precede substantial functional decline. Early recognition can alter the clinical trajectory by enabling referral before irreversible damage, appropriate monitoring intervals, or timely treatment to prevent vision loss. Conventional interpretation, however, is limited by specialist availability, interobserver variability, inconsistent image quality, and the growing volume of screening examinations. These limitations are especially evident in diabetes programs, elderly populations, rural healthcare centres, and health systems where retinal specialists are concentrated within urban centres.

Contemporary artificial intelligence systems employ convolutional neural networks, transformers, self-supervised learning, segmentation networks, and ensemble models to locate patterns in color fundus photographs and cross-sectional retinal scans. Their utility goes beyond disease labeling; modern systems can reject ungradable images, highlight suspected lesions, estimate disease severity, integrate bilateral findings, compare sequential examinations, and

recommend referral pathways. This expanded functionality represents a shift from algorithmic classification to comprehensive clinical decision support. The reference review similarly characterizes artificial intelligence as a technology that supports diagnosis, workflow optimization, teleophthalmology, personalized care, and governance, with diabetic retinopathy, age-related macular degeneration, and glaucoma highlighted as major applications [1].

However, high test-set accuracy does not guarantee safer or earlier clinical care. The model's performance may decline when images are acquired with different cameras, when disease prevalence shifts, when media opacity is prevalent, or when the evaluated population differs from the training cohort. Models may also exhibit overconfidence, rely on spurious image features, or generate referral recommendations that are not aligned with local capacity. Effective clinical integration therefore calls for careful attention to calibration, explainability, external validation, equity, and human monitoring. These factors are increasingly emphasized in high-impact journals, which require transparent methods, clinically relevant outcomes, and explicit discussion of generalizability.

This review intends to critically synthesize evidence from 2020 to 2025 regarding artificial intelligence-assisted interpretation of ophthalmic images for early detection of retinal disease. The specific objectives are to compare the diagnostic contributions of major imaging modalities; evaluate disease-specific applications in diabetic retinopathy, age-related macular degeneration, glaucoma, retinal vascular occlusion, and inherited retinal degeneration; assess the impact of image-quality control, explainability, and uncertainty on patient safety; and identify methodological priorities for translating promising models into routine screening and referral pathways. Unlike broader models, this review narrows its focus to early retinal detection and organizes evidence around the complete interpretation pathway, rather than encompassing treatment and ophthalmic surgery.

II. REVIEW METHODOLOGY

A structured narrative review design was selected because the topic encompasses heterogeneous imaging modalities, algorithm families, validation designs, clinical settings, and outcome measures. The review followed principles of transparent evidence synthesis without claiming a formal meta-analysis. Relevant literature was identified through structured searches of PubMed, Scopus, Web of Science Core Collection, IEEE Xplore, and ScienceDirect for peer-reviewed English-language publications issued between January 2020 and December 2025. Search concepts combined artificial intelligence, machine learning, deep learning, convolutional neural network, transformer, retinal image, fundus photography, optical coherence tomography, optical coherence tomography angiography, diabetic retinopathy, age-related macular degeneration, glaucoma, retinal vein occlusion, retinitis pigmentosa, screening, early detection, explainability, validation, and teleophthalmology. Boolean operators AND and OR were used to connect artificial intelligence methods with diseases, imaging modalities, and implementation outcomes.

Eligible publications were peer-reviewed English-language studies or substantial reviews published between January 2020 and December 2025. Included evidence addressed image-based detection, classification, segmentation, progression prediction, quality assessment, referral, or clinical implementation for retinal or optic nerve disease. Studies were prioritized when they reported external validation, multicenter evaluation, prospective testing, field deployment, subgroup analysis, or comparison with clinicians. Articles focused only on anterior-segment disease, non-imaging administrative automation, robotic surgery, or purely technical image reconstruction without diagnostic relevance were excluded. Foundational studies published before 2020 were excluded from the final reference list to maintain the requested 2020-2025 citation window; their historical contribution is acknowledged indirectly through recent reviews.

Information was extracted into a concept matrix covering population, disease, modality, algorithmic

task, reference standard, validation setting, reported performance, explainability, limitations, and clinical use. Because performance measures were not directly comparable across datasets and thresholds, findings were synthesized qualitatively rather than pooled. Particular attention was given to sensitivity for referable disease, specificity, area under the receiver operating characteristic curve, calibration, image gradability, external transportability, and operational outcomes such as screening completion or referral uptake. The methodological interpretation also considered whether the model was used as an autonomous system, an assistive reader, a triage tool, or a longitudinal monitoring aid.

Evidence quality was judged pragmatically using four questions. First, was the clinical target clearly defined and supported by an appropriate reference standard? Second, were training and evaluation data separated at the patient level and tested outside the development environment? Third, did the report describe image acquisition, missing data, class imbalance, and subgroup composition? Fourth, did the study evaluate consequences that matter to patients or services, rather than reporting discrimination alone? This system enables a balanced review because an algorithm with marginally lower accuracy may be more valuable when it is calibrated, transparent, robust to ungradable images, and integrated into a workable referral process.

Table 1. Review framework and eligibility criteria

Domain	Included	Excluded / deprioritized
Publication period	January 2020-December 2025	Pre-2020 references in the final bibliography
Evidence type	Peer-reviewed original studies and substantial reviews	Editorials, promotional claims, non-peer-reviewed summaries
Clinical scope	Early detection, classification, segmentation, progression, referral, implementation	Pure image reconstruction without diagnostic relevance
Imaging	Fundus, OCT, OCT-A, ultra-widefield, multimodal retinal data	Anterior-segment-only imaging
Priority features	External validation, prospective or real-world testing, subgroup reporting	Single-center development results without clinical context

III. OPHTHALMIC IMAGING AS THE SUBSTRATE FOR AI

Color fundus photography is the most established imaging substrate for scalable retinal artificial intelligence applications. It captures the optic disc, macula, vessels, hemorrhages, exudates, and pigmentary changes in a single en face view. Its principal advantages are speed, portability, familiarity, and compatibility with screening pathways. Smartphone adapters and portable cameras further extend acquisition beyond specialist clinics, which supports teleophthalmology and community programs. However, two-dimensional photographs can conceal depth-dependent pathology, and diagnostic reliability is reduced by small pupils, cataract, defocus, uneven illumination, and limited field of view. An effective artificial intelligence pipeline must therefore assess gradability before interpreting disease.

Optical coherence tomography provides micrometer-scale cross-sectional information and is especially useful for detecting intraretinal fluid, subretinal fluid, pigment epithelial detachment, drusen-associated changes, macular thinning, and retinal nerve fiber layer loss. Artificial intelligence can classify scans, segment retinal layers, quantify fluid, and compare repeated examinations. For early age-related macular degeneration and glaucoma, this structural information may reveal changes not obvious in a fundus photograph. The disadvantage is that devices differ in scan geometry, signal processing, segmentation software, and image export, creating

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722392

substantial domain shift. Models developed on a single manufacturer may therefore require adaptation or external recalibration before deployment elsewhere.

Optical coherence tomography angiography adds noninvasive vascular maps and can identify capillary dropout, neovascular networks, and alterations in the foveal avascular zone. Its promise is strongest when combined with structural optical coherence tomography rather than interpreted alone. Ultra-widefield photography also expands lesion coverage by visualizing the peripheral retina, which may improve recognition of diabetic lesions and vascular occlusion outside conventional fields. Yet peripheral distortion and variable eyelid or lash artifacts need tailored preprocessing.

Multimodal models attempt to combine complementary evidence from fundus photographs, optical coherence tomography, angiography, demographics, and clinical history. Such models can potentially distinguish visually similar conditions and estimate progression risk more effectively than single-image systems. Their practical challenge is missingness: not every patient has every imaging modality, and multimodal systems must remain reliable when one input is absent or poor. The most clinically useful architecture may therefore be modular, accepting available evidence while showing how confidence changes when data are incomplete. Figure 1 summarizes this interpretation pathway, from acquisition and quality assurance to explanation and clinical action.

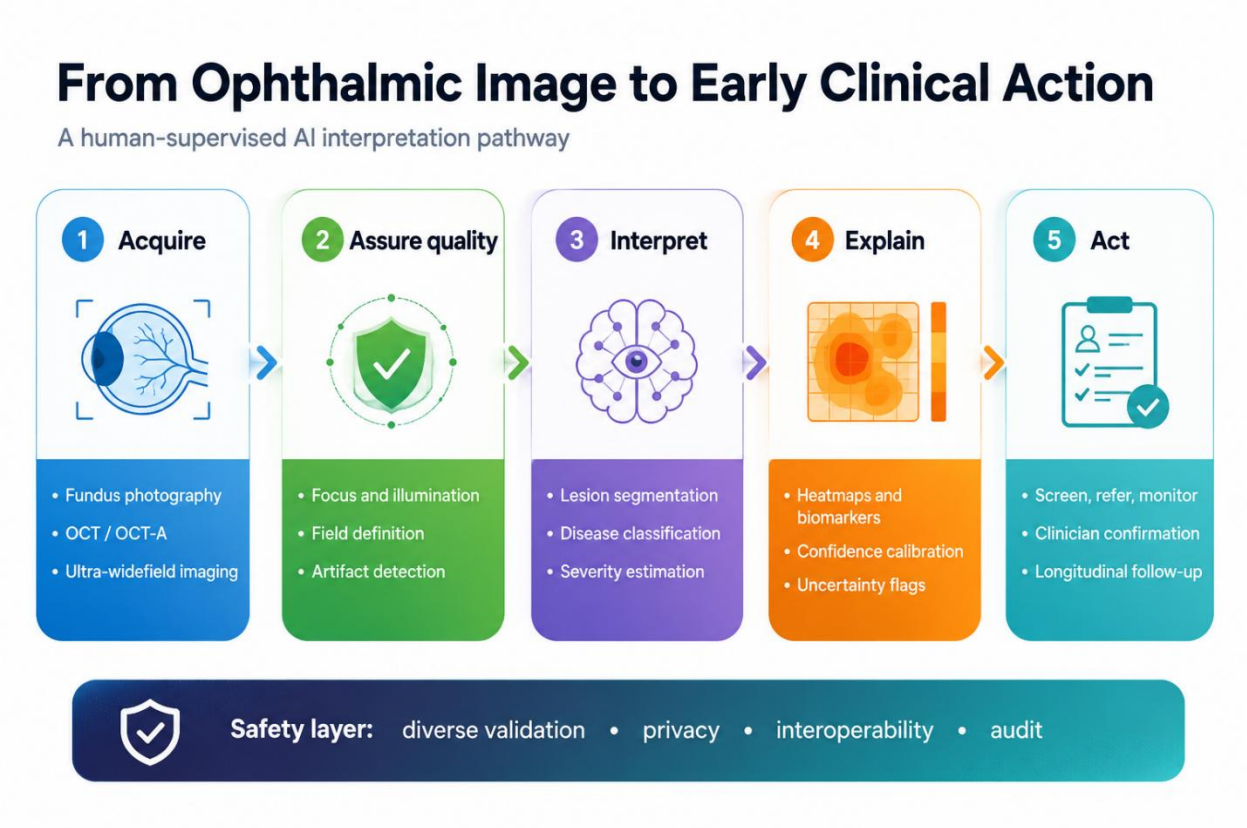

Figure 1. Human-supervised AI interpretation pathway for early retinal disease detection.

IV. DISEASE-SPECIFIC APPLICATIONS

Diabetic retinopathy is the most mature use case for autonomous or semi-autonomous retinal image interpretation. Deep learning systems detect microaneurysms, hemorrhages, exudates, cotton-wool spots, venous abnormalities, and proliferative changes, then map these data to referral thresholds. Recent real-world systematic evidence shows that automated screening can reach clinically useful

diagnostic performance, although results vary by device, population, and reference grading method [6]. Randomized evidence in young people with diabetes also indicates that autonomous artificial intelligence can increase screening completion and follow-up, demonstrating that workflow design can influence outcomes beyond diagnostic accuracy [4]. The strongest implementations combine immediate results with a clear referral process; otherwise, detection may not translate into treatment.

Age-related macular degeneration requires recognition of drusen, pigmentary abnormalities, geographic atrophy, and exudative activity. Fundus-based models are suitable for broad stage classification, while optical coherence tomography is more informative for early exudative change, fluid, and treatment monitoring. Recent systematic work highlights the predictive capability of artificial intelligence-based optical coherence tomography analysis for progression, but also identifies inconsistent endpoints and limited external validation [17]. Multimodal systems may improve differentiation between late dry and neovascular disease by integrating biomarkers across imaging types [18]. For early detection, the clinically relevant question is not simply whether age-related macular degeneration is present, but whether the observed pattern justifies shorter surveillance, specialist review, or urgent anti-vascular endothelial growth factor assessment.

Glaucoma presents a different problem because structural and functional change may be gradual, asymmetric, and confounded by normal anatomical variation. Artificial intelligence can assess optic disc photographs, retinal nerve fiber layer maps, ganglion cell measurements, and visual fields. Reviews published since 2020 describe progress in glaucoma diagnosis and progression detection while stressing the need for translation into routine care [9,11]. Longitudinal models are especially valuable because early glaucoma may be uncertain on a single examination. Systems that compare repeated scans, quantify change, and combine structural with functional evidence are more clinically credible than static classifiers. However, myopia, tilted discs, segmentation errors, and device-specific normative databases can generate false alarms.

Retinal vascular occlusion is increasingly studied using color fundus photography, fluorescein angiography, optical coherence tomography, and optical coherence tomography angiography. Artificial intelligence can classify branch and central occlusion, delineate ischemic areas, and detect macular edema. Multicenter research on fluorescein angiography demonstrates the feasibility of automated interpretation, but the invasiveness and variable timing of angiography limit population screening [26]. Fundus and optical coherence tomography models may therefore be more practical for triage, with angiographic artificial intelligence reserved for detailed specialist assessment.

Inherited retinal degeneration, including retinitis pigmentosa, illustrates the value of combining imaging with genotype and phenotype information. Early signs may be subtle and heterogeneous, and rare-disease datasets are typically small. Artificial intelligence-assisted early detection has been demonstrated using retinal imaging, but robust deployment requires multicenter data sharing and careful handling of genetic privacy [24]. Transfer learning, self-supervised pretraining, and federated collaboration may partly address sample scarcity. Across all disease groups, Figure 2 illustrates that modality value is task dependent: fundus photography allows scalable screening, optical coherence tomography supports structural detection, angiographic imaging supports vascular characterization, and multimodal records support complex risk estimation.

Table 2. Disease-specific AI interpretation priorities

Disease group	Primary images	Early AI target	Clinical value	Key limitation
Diabetic retinopathy	Color fundus; ultra-widefield; OCT	Referable disease and macular edema	Scalable screening and immediate referral	Ungradable images and population shift

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722392

Age-related macular degeneration	OCT; fundus; multimodal	Drusen, fluid, atrophy, progression risk	Earlier surveillance or treatment assessment	Endpoint heterogeneity and device shift
Glaucoma	Disc photos; OCT; visual fields	Optic neuropathy and longitudinal progression	Earlier risk stratification	Myopia, anatomy, segmentation error
Retinal vascular occlusion	Fundus; OCT; OCT-A; angiography	Occlusion type, ischemia, edema	Triage and complication detection	Smaller datasets and modality timing
Inherited retinal degeneration	Fundus; autofluorescence; OCT; genotype	Subtle phenotype and progression pattern	Earlier specialist/genetic pathway	Rare classes and privacy

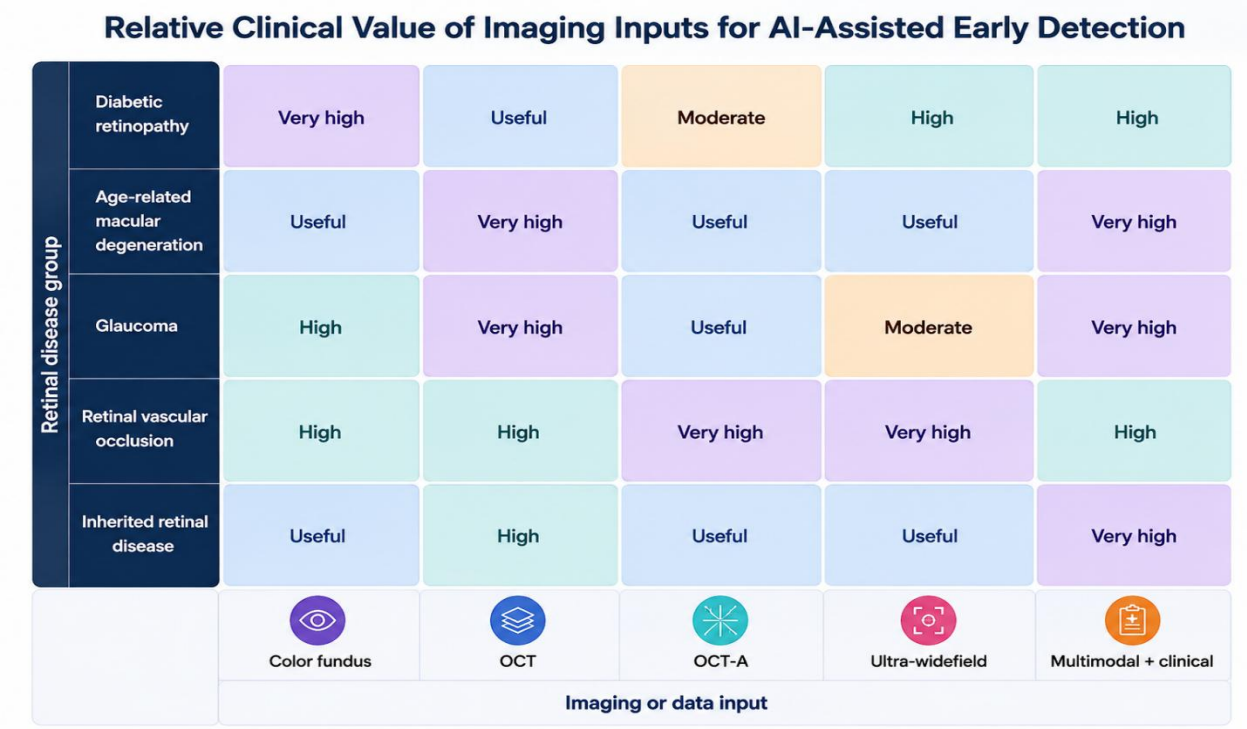

Figure 2. Conceptual comparison of imaging inputs for AI-assisted early detection across retinal disease groups.

V. FROM CLASSIFICATION TO CLINICALLY MEANINGFUL INTERPRETATION

The clinical usefulness of artificial intelligence depends on more than a binary output. A credible system should indicate whether an image is gradable, what anatomical region influenced the decision, how confident the model is, and what action is recommended. Image-quality assessment is therefore the first diagnostic step. Quality models can identify blur, underexposure, overexposure, decentration, eyelid obstruction, motion artifact, and incomplete scan coverage. Rejecting poor images may reduce false reassurance, but excessive rejection may disadvantage older patients and those with cataract. Services must monitor rejection rates by age, ethnicity, device, and location to ensure that quality control does not become an unintended barrier.

Lesion localization increases interpretability by showing microaneurysms, hemorrhages, drusen, fluid compartments, optic disc boundaries, or nerve fiber layer defects. Segmentation-based outputs can be

reviewed by clinicians and can support longitudinal quantification. Yet saliency maps alone should not be treated as proof that a model reasons correctly. Some visual explanations are unstable, coarse, or influenced by preprocessing. More useful explanations combine localized findings with structured statements, confidence intervals, image-quality warnings, and comparison with prior examinations.

Calibration is equally important. A model can have a high area under the curve while producing probabilities that do not correspond to real risk. Poor calibration complicates threshold selection and may overload referral clinics. Thresholds should reflect disease severity, consequences of a missed case, local specialist capacity, and whether artificial intelligence is used for screening or diagnosis. For sight-threatening diabetic retinopathy, high sensitivity is generally prioritized; for a specialist triage tool, balanced specificity may be necessary to prevent unnecessary urgent referrals.

Uncertainty-aware systems can flag unfamiliar or ambiguous images rather than forcing a confident label. This is especially important for rare diseases, coexisting pathology, postoperative eyes, unusual pigmentation, and images from unrepresented devices. An abstention pathway should route uncertain cases to a human grader. Such human-artificial intelligence collaboration is not a weakness; it is a safety architecture. The clinician remains responsible for integrating symptoms, visual acuity, medical history, and examination findings that may not be represented in the image.

Longitudinal interpretation is a major frontier. Early disease is often defined by change rather than a single abnormality. Algorithms that register sequential images, quantify structural differences, and model progression can support earlier intervention. However, apparent change may result from scan placement, device upgrade, segmentation drift, or image-quality variation. Longitudinal systems therefore need harmonization, repeatability analysis, and explicit separation of biological progression from technical noise.

VI. CLINICAL TRANSLATION AND WORKFLOW INTEGRATION

Artificial intelligence can enter practice in four principal roles: autonomous screening, assistive reading, triage, and monitoring. Autonomous screening produces a result without specialist interpretation and is appropriate only when the intended population, disease threshold, device, and referral pathway have been prospectively validated. Assistive reading presents model findings to a clinician, possibly boosting sensitivity or speed while preserving clinical oversight. Triage prioritizes examinations for rapid review, and monitoring compares serial images to identify possible progression. Each role needs a different evaluation design and a different balance among sensitivity, specificity, and explainability.

Teleophthalmology is a natural deployment environment because images can be captured in primary care or community settings and interpreted centrally. Recent reviews show that combining telemedicine with artificial intelligence may expand diabetic retinopathy screening, reduce travel, and support underserved populations [7,8]. Smartphone-enabled imaging broadens access further, although variable optics, compression, and operator training can affect reliability [28]. A safe service should include standardized acquisition protocols, automated quality feedback, secure transmission, rapid referral for positive cases, and tracking of whether referred patients attend.

Interoperability determines whether an algorithm becomes part of care or remains an isolated demonstration. Outputs should integrate with electronic health records, picture archiving systems, and scheduling platforms. Structured reports should state the examined eye, image type, quality, predicted disease, severity, confidence, and recommended next step. Audit records should record model version, threshold, input identifiers, and clinician action. These records are essential when methods are updated or when performance is investigated.

Economic value must also be measured in context. Artificial intelligence may reduce manual grading,

but savings can be offset by camera purchase, software licensing, data integration, cybersecurity, quality assurance, and increased referrals. Evidence from real-world diabetic retinopathy screening indicates that diagnostic accuracy and cost-effectiveness are related but are not interchangeable [6]. A lower-cost system that generates many false positives may burden specialist services, while a more expensive but better calibrated system may preserve capacity. Decision makers therefore need analyses that include downstream testing, treatment, patient travel, missed disease, and staff time.

Implementation should proceed in stages: silent retrospective evaluation, prospective shadow testing, controlled clinician-assisted use, and monitored routine deployment. During shadow testing, the model functions without influencing care, allowing comparison with local practice and identification of unexpected failure modes. After deployment, performance should be monitored for drift, subgroup disparities, changes in prevalence, and changes in imaging hardware. Governance committees should include ophthalmologists, imaging staff, data scientists, information-security specialists, service managers, and patient representatives.

VII. BIAS, PRIVACY, REGULATION, AND ACCOUNTABILITY

Bias can enter through image acquisition, dataset composition, labeling, model design, and deployment. Retinal pigmentation, age, media opacity, myopia, comorbidity, and camera type may influence image appearance. If training data overrepresent one population or device, performance can be lower elsewhere. Ethical analyses in ophthalmology therefore emphasize representative data, subgroup reporting, transparency, and sustained monitoring [29]. Fairness should be evaluated using clinically meaningful error rates, particularly false negatives for referable disease, rather than a single overall metric.

Reference labels can also be biased. Human graders disagree, and consensus labels may conceal uncertainty. Studies should report grader expertise, adjudication procedures, and the diagnostic criteria used. Where possible, labels should incorporate follow-up, multimodal confirmation, or treatment decisions rather than relying on one image. Models developed on imperfect labels may reproduce institutional habits instead of biological truth.

Privacy is central because retinal images can be linked with demographic, metabolic, genetic, and longitudinal health data. Data protection requires encryption, access control, de-identification, retention policies, and documented consent or lawful processing. Ophthalmic privacy scholarship warns that image-rich artificial intelligence systems create risks when datasets are shared across institutions or commercial partners [30]. Federated learning may reduce movement of raw data, but it does not eliminate risks from model updates, re-identification, or unequal participation.

Regulatory evaluation ought to match the intended use. A screening device that independently determines referral carries different risks from software that merely highlights suspicious regions for a specialist. Developers should define the target population, input devices, contraindications, failure responses, and human role. Significant model updates may alter performance and require renewed validation. Post-market surveillance is particularly important because real-world prevalence and imaging conditions change.

Accountability must remain explicit. Patients should know when artificial intelligence contributes to interpretation, and clinicians should comprehend its limitations. The system should not obscure responsibility behind a likelihood estimate. Institutions need procedures for reporting errors, reviewing adverse events, and suspending use when drift or safety concerns emerge. Transparent governance protects patients and also supports clinician trust, which is essential for adoption.

VIII. RESEARCH GAPS AND FUTURE DIRECTIONS

Future research needs to move from retrospective accuracy contests toward prospective, outcome-centered evaluation. The most meaningful endpoint is

not whether a model reproduces a grader label, but whether it enables earlier diagnosis, appropriate referral, timely treatment, preserved vision, and equitable access. Cluster-randomized or stepped-wedge studies can compare artificial intelligence-supported pathways with usual care while accounting for service-level effects. Trials should report screening uptake, ungradable rates, referral completion, waiting time, treatment initiation, visual outcomes, and patient experience.

External validation must become routine. Multicenter datasets should include different cameras, scan protocols, ethnic groups, ages, and levels of disease prevalence. Performance should be reported separately for important subgroups and for coexisting pathology. Temporal validation is also necessary because devices, clinical practice, and population risk change. Studies should publish calibration curves, decision-curve analyses, and threshold-specific consequences rather than only area under the curve.

Self-supervised and foundation models may improve data efficiency by learning general retinal representations from large unlabeled image collections. Their broad capability, however, increases the requirement for careful task-specific validation. Multimodal models that combine images with visual acuity, laboratory values, genetics, and clinical history may improve risk prediction, but missing data, privacy, and explainability must be addressed. Models should be able to state when an absent modality reduces confidence.

Synthetic data may help balance rare disease classes, yet generated images can introduce unrealistic patterns or leak information from training data. Their use should be transparently reported and validated by both technical and clinical experts. Privacy-preserving methods, including federated learning and secure computation, should be evaluated not only for accuracy but also for governance, participation equity, and computing overhead.

Explainable artificial intelligence research should give priority to explanations that change clinical decisions safely. Structured lesion summaries, uncertainty estimates, and longitudinal comparisons are likely to be more useful than decorative heatmaps. Human-factors studies should examine how clinicians respond to correct and incorrect suggestions, whether automation bias occurs, and how interface design influences attention. Training should include examples of model failure, not only demonstrations of success.

Finally, implementation science should examine how artificial intelligence affects different health systems. A model that is efficient in a tertiary center may fail in a rural screening program with intermittent connectivity and limited referral capacity. Co-design with patients and frontline staff can identify practical barriers early. International collaboration is needed to build representative evidence while observing local control of data. These priorities will help transform artificial intelligence from a promising image classifier into a reliable component of preventive retinal care.

A further priority is benchmark governance. Public benchmarks should document licensing, image provenance, duplicate detection, population makeup, acquisition devices, labeling uncertainty, and permitted uses. Hidden test sets can reduce overfitting, while periodic benchmark renewal can discourage optimization to stale datasets. Reporting checklists should require patient-level splits, confidence intervals, calibration, external validation, and a description of clinical consequences at the selected threshold. Open code and model cards can improve reproducibility, although privacy and intellectual-property constraints may limit complete release. Independent evaluation laboratories could test commercial systems under standardized but clinically realistic conditions. These laboratories should include degraded images, uncommon co-pathology, different pigmentation, pediatric and older eyes, and devices not represented during development. Procurement contracts should require access to performance reports, incident procedures, version histories, and data-use terms. This governance infrastructure would make comparisons more meaningful and would help health systems select tools according to local needs rather than marketing claims.

A practical research agenda should also define minimum evidence for deployment. Before routine use, a model should demonstrate stable performance on local images, clinically acceptable sensitivity at the intended threshold, calibrated probabilities, and a safe response to ungradable or unfamiliar inputs. Prospective evaluation should confirm that staff understand the output and that positive results lead to completed referrals. After implementation, dashboards should track image rejection, false-negative review findings, referral volume, attendance, subgroup performance, and changes associated with software updates. These measures should be reviewed at predefined intervals by a multidisciplinary governance group. When deterioration is detected, the organization should be able to pause the system, restore a previous version, or return to human grading without interrupting care. Researchers should additionally distinguish technical validation from clinical validation and implementation evaluation. Technical validation asks whether predictions are accurate; clinical validation asks whether they are useful for the intended patients; implementation evaluation asks whether the service can deliver the promised benefit consistently. Treating these stages as separate but connected will reduce premature adoption. It will also encourage authors to describe workflow dependencies, staffing assumptions, referral capacity, and contingency plans. Such reporting would make studies easier to compare, support responsible procurement, and provide journal reviewers with clearer evidence that an innovation is clinically mature rather than merely computationally impressive. Patient involvement should be documented from protocol design through interface testing, because acceptability, consent, accessibility, and trust can determine whether technically excellent screening reaches the people at greatest risk of preventable blindness in everyday practice and produces timely, equitable clinical benefit.

IX. LIMITATIONS OF THIS REVIEW

This review has limitations. It uses a structured narrative method rather than a registered systematic review and does not pool diagnostic performance estimates. The 2020-2025 date restriction supports contemporary relevance yet excludes fundamental studies that established many current architectures. English-language selection may underrepresent evidence from rapidly developing programs in non-English-speaking regions. The literature is also affected by publication bias, inconsistent reporting, overlapping datasets, and rapidly changing model versions. Consequently, the review emphasizes patterns of evidence and translation requirements rather than ranking individual commercial systems.

X. CONCLUSION

Artificial intelligence-assisted interpretation of ophthalmic images can strengthen early detection of retinal disease when it is embedded within a complete, human-supervised care pathway. Fundus photography provides scalable screening, optical coherence tomography reveals structural biomarkers, angiographic methods characterize vascular pathology, and multimodal models may improve individualized risk estimation. The strongest evidence is currently found in diabetic retinopathy, while age-related macular degeneration and glaucoma increasingly benefit from optical coherence tomography and longitudinal analysis. Reliable translation requires image-quality control, calibrated thresholds, uncertainty handling, external validation, equitable subgroup performance, secure data governance, interoperable reporting, and explicit clinician accountability.

The central lesson is that diagnostic performance alone is insufficient. A clinically valuable system must improve the timing and appropriateness of action without creating unsafe reassurance, unmanageable referrals, or new disparities. Future studies should therefore measure patient and service outcomes, compare human-artificial intelligence workflows prospectively, and report failure modes transparently. Used in this disciplined manner, artificial intelligence can extend specialist expertise, support teleophthalmology, and help prevent avoidable vision loss while safeguarding the essential role of clinical judgment.

REFERENCES

[1] Olawade, D.B.; Weerasinghe, K.; Mathugamage, M.D.D.E.; Odetayo, A.; Aderinto, N.; Teke, J.; Boussios, S. Enhancing Ophthalmic Diagnosis and Treatment with Artificial Intelligence. Medicina 2025, 61(3), 433. https://doi.org/10.3390/medicina61030433.

[2] Tan, T.F.; Thirunavukarasu, A.J.; Jin, L.; et al. Artificial Intelligence and Digital Health in Global Eye Health: Opportunities and Challenges. The Lancet Global Health 2023, 11(9), e1432-e1443. https://doi.org/10.1016/S2214-109X(23)00323-6.

[3] Kropp, M.; Golubnitschaja, O.; Mazurakova, A.; et al. Diabetic Retinopathy as the Leading Cause of Blindness and Early Predictor of Cascading Complications-Risks and Mitigation. EPMA Journal 2023, 14(1), 21-42. https://doi.org/10.1007/s13167-023-00314-8.

[4] Wolf, R.M.; Channa, R.; Liu, T.Y.A.; et al. Autonomous Artificial Intelligence Increases Screening and Follow-Up for Diabetic Retinopathy in Youth: The ACCESS Randomized Controlled Trial. Nature Communications 2024, 15, 421. https://doi.org/10.1038/s41467-023-44676-z.

[5] Campbell, J.P.; Mathenge, C.; Cherwek, H.; Balaskas, K.; Pasquale, L.R.; Keane, P.A.; Chiang, M.F. Artificial Intelligence to Reduce Ocular Health Disparities: Moving from Concept to Implementation. Translational Vision Science & Technology 2021, 10(3), 19. https://doi.org/10.1167/tvst.10.3.19.

[6] Joseph, S.; Selvaraj, J.; Mani, I.; Kumaragurupari, T.; Shang, X.; Mudgil, P.; Ravilla, T.; He, M. Diagnostic Accuracy of Artificial Intelligence-Based Automated Diabetic Retinopathy Screening in Real-World Settings: A Systematic Review and Meta-Analysis. American Journal of Ophthalmology 2024, 263, 214-230. https://doi.org/10.1016/j.ajo.2024.02.012.

[7] Chokshi, T.; Cruz, M.J.; Ross, J.; Yiu, G. Advances in Teleophthalmology and Artificial Intelligence for Diabetic Retinopathy Screening: A Narrative Review. Annals of Eye Science 2024, 9, 9. https://doi.org/10.21037/aes-23-10.

[8] Pieczynski, J.; Kuklo, P.; Grzybowski, A. The Role of Telemedicine, In-Home Testing and Artificial Intelligence to Alleviate an Increasingly Burdened Healthcare System: Diabetic Retinopathy. Ophthalmology and Therapy 2021, 10, 445-464. https://doi.org/10.1007/s40123-021-00353-2.

[9] Mursch-Edlmayr, A.S.; Ng, W.S.; Diniz-Filho, A.; Sousa, D.C.; Arnould, L.; et al. Artificial Intelligence Algorithms to Diagnose Glaucoma and Detect Glaucoma Progression: Translation to Clinical Practice. Translational Vision Science & Technology 2020, 9(2), 55. https://doi.org/10.1167/tvst.9.2.55.

[10] Zhu, Y.; Salowe, R.; Chow, C.; et al. Advancing Glaucoma Care: Integrating Artificial Intelligence in Diagnosis, Management, and Progression Detection. Bioengineering 2024, 11(2), 122. https://doi.org/10.3390/bioengineering11020122.

[11] Yousefi, S. Clinical Applications of Artificial Intelligence in Glaucoma. Journal of Ophthalmic & Vision Research 2023, 18(1), 97-112. https://doi.org/10.18502/jovr.v18i1.12730.

[12] Zhang, L.; Tang, L.; Xia, M.; Cao, G. The Application of Artificial Intelligence in Glaucoma Diagnosis and Prediction. Frontiers in Cell and Developmental Biology 2023, 11, 1173094. https://doi.org/10.3389/fcell.2023.1173094.

[13] Hussain, S.; Chua, J.; Wong, D.; et al. Predicting Glaucoma Progression Using a Deep Learning Framework Guided by a Generative Algorithm. Scientific Reports 2023, 13, 19960. https://doi.org/10.1038/s41598-023-46253-2.

[14] Bhartiya, S. Glaucoma Screening: Is AI the Answer? Journal of Current Glaucoma Practice 2022, 16(2), 71-73. https://doi.org/10.5005/jp-journals-10078-1380.

[15] Fleckenstein, M.; Schmitz-Valckenberg, S.; Chakravarthy, U. Age-Related Macular Degeneration: A Review. JAMA 2024, 331(2), 147-157. https://doi.org/10.1001/jama.2023.26074.

[16] Elsharkawy, M.; Elrazzaz, M.; Ghazal, M.; et al. Role of Optical Coherence Tomography Imaging in Predicting Progression of Age-Related Macular Disease: A Survey. Diagnostics 2021, 11(12), 2313. https://doi.org/10.3390/diagnostics11122313.

[17] Muntean, G.A.; Marginean, A.; Groza, A.; et al. The Predictive Capabilities of Artificial Intelligence-Based OCT Analysis for Age-Related Macular Degeneration Progression: A Systematic Review. Diagnostics 2023, 13(14), 2464. https://doi.org/10.3390/diagnostics13142464.

[18] Thakoor, K.A.; Yao, J.; Bordbar, D.; et al. A Multimodal Deep Learning System to Distinguish Late Stages of AMD and to Compare Expert versus AI Ocular Biomarkers. Scientific Reports 2022, 12, 2585. https://doi.org/10.1038/s41598-022-06273-w.

[19] Mares, V.; Nehemy, M.B.; Bogunovic, H.; et al. AI-Based Support for Optical Coherence Tomography in Age-Related Macular Degeneration. International Journal of Retina and Vitreous 2024, 10, 31. https://doi.org/10.1186/s40942-024-00549-1.

[20] Saleh, G.A.; Batouty, N.M.; Haggag, S.; et al. The Role of Medical Image Modalities and AI in the Early Detection, Diagnosis and Grading of Retinal Diseases: A Survey. Bioengineering 2022, 9(8), 366. https://doi.org/10.3390/bioengineering9080366.

[21] Liu, X.; Zhao, C.; Wang, L.; Wang, G.; Lv, B.; Lv, C.; Xie, J.; Wang, N. Evaluation of an OCT-AI-Based Telemedicine Platform for Retinal Disease Screening and Referral in a Primary Care Setting. Translational Vision Science & Technology 2022, 11(3), 4. https://doi.org/10.1167/tvst.11.3.4.

[22] Mantena, S.; Celi, L.A.; Keshavjee, S.; Beratarrechea, A. Improving Community Health-Care Screenings with Smartphone-Based AI Technologies. The Lancet Digital Health 2021, 3(5), e280-e282. https://doi.org/10.1016/S2589-7500(21)00054-6.

[23] Parmar, U.P.S.; Surico, P.L.; Singh, R.B.; et al. Artificial Intelligence for Early Diagnosis of Retinal Diseases. Medicina 2024, 60(4), 527. https://doi.org/10.3390/medicina60040527.

[24] Chen, T.C.; Lim, W.S.; Wang, V.Y.; et al. Artificial Intelligence-Assisted Early Detection of Retinitis Pigmentosa-The Most Common Inherited Retinal Degeneration. Journal of Digital Imaging 2021, 34(4), 948-958. https://doi.org/10.1007/s10278-021-00479-6.

[25] Ji, Y.K.; Hua, R.R.; Liu, S.; et al. Intelligent Diagnosis of Retinal Vein Occlusion Based on Color Fundus Photographs. International Journal of Ophthalmology 2024, 17(1), 1-6. https://doi.org/10.18240/ijo.2024.01.01.

[26] Huang, S.; Jin, K.; Gao, Z.; et al. Automated Interpretation of Retinal Vein Occlusion Based on Fundus Fluorescein Angiography Images Using Deep Learning: A Retrospective, Multicenter Study. Heliyon 2024, 10(13), e33108. https://doi.org/10.1016/j.heliyon.2024.e33108.

[27] Li, Z.; Wang, L.; Wu, X.; et al. Artificial Intelligence in Ophthalmology: The Path to the Real-World Clinic. Cell Reports Medicine 2023, 4(7), 101095. https://doi.org/10.1016/j.xcrm.2023.101095.

[28] Jin, K.; Li, Y.; Wu, H.; et al. Integration of Smartphone Technology and Artificial Intelligence for Advanced Ophthalmic Care: A Systematic Review. Advances in Ophthalmology Practice and Research 2024, 4(3), 120-127. https://doi.org/10.1016/j.aopr.2024.03.003.

[29] Evans, N.G.; Wenner, D.M.; Cohen, I.G.; et al. Emerging Ethical Considerations for the Use of Artificial Intelligence in Ophthalmology. Ophthalmology Science 2022, 2(2), 100141. https://doi.org/10.1016/j.xops.2022.100141.

[30] Tom, E.; Keane, P.A.; Blazes, M.; Pasquale, L.R.; Chiang, M.F.; Lee, A.Y.; Lee, C.S. Protecting Data Privacy in the Age of AI-Enabled Ophthalmology. Translational Vision Science & Technology 2020, 9(2), 36. https://doi.org/10.1167/tvst.9.2.36.