Ophthalmol Ther (2026) 15:569–589
https://doi.org/10.1007/s40123-026-01322-3

REVIEW

Artificial Intelligence-Based Medical Devices for Diabetic Retinopathy Screening in the European Union

Andrzej Grzybowski · Kai Jin

Received: November 14, 2025 / Accepted: January 14, 2026 / Published online: January 30, 2026

ABSTRACT

Background: Diabetic retinopathy (DR) remains a leading cause of preventable blindness, yet screening programs across Europe face persistent workforce and capacity constraints amid rising diabetes prevalence. Artificial intelligence (AI)-enabled screening platforms have been developed to support scalable DR detection; however, their regulatory status, validation approaches, and implementation readiness vary considerably.

Methods: We conducted a targeted scoping review of 13 CE-certified AI systems for autonomous or semi-autonomous DR detection available in the European Union as of October 23, 2025 (IDx-DR, EyeArt, RetCAD, Mona DR, Retmarker DR, SELENA+, Remidio Medios AI, RetinoScan, Aireen DR, OphthAI, LuxIA, Airdoc-Eye DR, and Vistel). Data were charted across predefined domains, including device designation, regulatory classification, evidence sources, validation study design, reported diagnostic performance metrics, and implementation-related considerations. The review aimed to map the extent and nature of available evidence without conducting quantitative synthesis or comparative ranking.

Results: Most systems employed deep-learning-based fundus image analysis, often incorporating automated image-quality assessment. Reported sensitivities and specificities for referable DR (RDR) varied across systems, generally falling within ranges consistent with regulatory expectations; however, reporting standards and study designs were heterogeneous, limiting direct comparison. Several systems were supported by multicenter or prospective evaluations, while others relied primarily on retrospective datasets. A subset of platforms reported multi-disease detection capabilities. Evidence specific to sight-threatening DR (STDR) was less

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40123-026-01322-3.

A. Grzybowski (✉)
Department of Ophthalmology, University of Warmia and Mazury, 10-719 Olsztyn, Poland
e-mail: ae.grzybowski@gmail.com

A. Grzybowski
Institute for Research in Ophthalmology, Foundation for Ophthalmology Development, 61-553 Poznan, Poland

K. Jin (✉)
Eye Center of Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310009, Zhejiang, China
e-mail: jinkai@zju.edu.cn

K. Jin
Zhejiang Provincial Key Laboratory of Ophthalmology, Zhejiang Provincial Clinical Research Center for Eye Diseases, Zhejiang Provincial Engineering Institute On Eye Diseases, Hangzhou 310009, Zhejiang, China

frequently described and demonstrated wider variability. Non-EU regulatory pathways were mentioned in some reports, but were outside the primary scope of this review. Other systems demonstrate high diagnostic accuracy in controlled evaluations, though performance for STDR remains limited (mean ≈ 80%), largely due to reliance on single-modality 2D fundus imaging without optical coherence tomography (OCT) integration. Implementation-related evidence, including workflow integration and monitoring requirements under the EU Medical Device Regulation (MDR), was limited across systems.

Conclusions: CE-certified AI systems for DR detection represent a diverse and rapidly evolving landscape. While substantial progress has been made in regulatory classification and validation efforts, evidence remains heterogeneous, particularly for STDR detection and real-world implementation. Future research should prioritize consistent reporting standards, evaluation of multimodal approaches, and studies addressing real-world effectiveness to support safe and equitable deployment under the evolving EU regulatory framework.

Keywords: Diabetic retinopathy; Artificial intelligence; CE certification; EU Medical Device Regulation; EU AI Act; Screening

Key Summary Points

Why carry out this study?

Diabetic retinopathy remains a leading cause of preventable blindness, yet screening programs in Europe continue to face workforce shortages and capacity limitations.

Multiple CE-certified AI systems for DR screening are now available, but their clinical evidence, regulatory designations, and real-world readiness vary considerably.

The study provides a scoping review synthesizing the regulatory status, validation evidence, and implementation considerations of currently CE-certified DR AI systems.

What was learned from the study?

CE-certified AI systems demonstrated high diagnostic accuracy for referable DR detection, with IDx-DR and EyeArt supported by the strongest prospective real-world evidence.

Sight-threatening DR detection performance remains moderate, reflecting limitations of single-modality fundus imaging and the need for multimodal integration (e.g., optical coherence tomography).

Future progress requires improved transparency of dataset provenance, strengthened post-market monitoring, and head-to-head comparative trials to ensure trustworthy, equitable real-world implementation under the EU AI Act.

INTRODUCTION

Diabetic retinopathy (DR), a chronic microvascular complication of diabetes mellitus, remains one of the foremost causes of preventable blindness globally, accounting for a substantial burden in both working-age and elderly populations [1, 2]. Despite robust evidence that timely detection and treatment, such as panretinal photocoagulation or intravitreal anti-VEGF therapy, can prevent more than 90% of vision loss, screening programs worldwide continue to face logistical and capacity-related challenges [3]. The exponential rise in diabetes prevalence, coupled with shortages of trained ophthalmic graders and unequal access to specialized care, has created a widening screening gap that threatens timely intervention.

Artificial intelligence (AI), particularly deep learning, based image analysis, has emerged as a transformative tool to bridge this gap by enabling automated, high-throughput, and standardized interpretation of retinal photographs [4]. Over the past decade, multiple AI systems have

been developed for DR detection, ranging from assistive decision-support tools to fully autonomous diagnostic algorithms capable of rendering clinical judgments without physician input [5, 6]. Early evaluations have reported encouraging diagnostic performance across diverse settings; however, the scope, methodology, and reporting standards of these studies vary, underscoring the need to better characterize the existing evidence base.

Within Europe, several of these AI systems have achieved CE certification, signifying regulatory approval under the EU Medical Device Regulation (MDR; Regulation (EU) 2017/745) [7]. CE certification indicates compliance with stringent requirements for safety, performance, and post-market oversight, yet publicly available documentation describing clinical validation, dataset characteristics, and implementation processes varies considerably across systems [8–10]. In particular, information on dataset provenance (e.g., geographic distribution, imaging devices, grading protocols, and management of ungradable images) is inconsistently reported, limiting assessment of generalizability and equity considerations. While recent technological innovations such as vision foundation models and multimodal imaging architectures have been reported in the literature, these developments remain outside the scope of currently CE-certified autonomous DR screening systems and are therefore referenced only to contextualize the broader evolution of the field.

Given these considerations, we conducted a scoping review to describe the characteristics of CE-certified AI systems for DR screening in the European Union. This review aims to map the existing evidence related to device architectures, regulatory designation, validation study designs, and reporting practices, and to identify the gaps relevant to transparency and post-market performance. Consistent with scoping review methodology, we synthesized data descriptively without statistical pooling or comparative performance assessment. This review therefore provides a narrative, non-comparative synthesis of CE-certified DR screening systems and their supporting evidence, framed within the regulatory context of the MDR and the forthcoming EU Artificial Intelligence Act.

Ethical Approval

This article is based on previously conducted studies and does not contain any new studies with human participants or animals performed by either author.

REGULATORY FRAMEWORK IN THE EUROPEAN UNION

The European regulatory framework for AI-enabled medical devices is primarily governed by the Medical Device Regulation (MDR), which replaced the earlier Medical Device Directive (MDD) in 2021 [11]. The MDR classifies software performing diagnostic or screening tasks as at least Class IIa, acknowledging their moderate to high clinical risk. Conformity assessment is undertaken by Notified Bodies, which evaluate technical documentation, clinical performance, data integrity, cybersecurity, and post-market surveillance (PMS) before CE marking can be granted.

The transition from MDD to MDR has strengthened requirements for clinical validation, emphasizing prospective, representative, and methodologically transparent evidence to demonstrate both safety and effectiveness in target populations. Furthermore, developers must implement lifecycle monitoring of algorithmic performance, traceability of training data, and PMS mechanisms such as periodic safety update reports (PSURs) [12]. These elements are critical for mitigating "model drift" and ensuring sustained reliability for market deployment.

The European Database on Medical Devices (EUDAMED) is a cornerstone of the MDR, designed to enhance transparency and traceability across the medical device lifecycle [13]. EUDAMED serves as a centralized platform for registering devices, economic operators (e.g., manufacturers, importers), and clinical investigations, ensuring compliance with MDR requirements. For AI-based medical devices, EUDAMED is expected to support PMS activities through the registration of vigilance reports and safety updates. Manufacturers must also use

EUDAMED to assign Unique Device Identification (UDI) codes, supporting traceability and regulatory oversight.

In parallel, the European Artificial Intelligence Act (AI Act)—expected to take effect in the coming years—will further delineate requirements for "high-risk" AI systems, including medical diagnostic software [14]. It introduces obligations for algorithmic transparency, interpretability, human oversight, continuous risk management, and data governance. The AI Act complements the MDR by addressing aspects specific to adaptive and data-driven systems, such as retraining controls and dataset governance. Together, these frameworks contribute to evolving expectations for transparency and oversight in medical AI [15, 16].

CE-CERTIFIED AI MEDICAL DEVICES FOR DIABETIC RETINOPATHY IN THE EU

As of 2025, a limited number of ophthalmic AI systems have obtained CE certification for autonomous or semi-autonomous DR detection (Table 1). These include IDx-DR (Digital Diagnostics, USA), EyeArt (Eyenuk, USA), RetCAD (Thirona, Netherlands), and Mona DR (MonA, Belgium), among others [8]. In accordance with the EU MDR, we restricted our analysis to Class IIa and IIb AI-based devices, as software performing diagnostic or screening functions is classified as at least Class IIa. All regulatory classifications and CE statuses reflect information available as of October 2025. Each is reported to rely on deep convolutional neural network (CNN) architectures trained on large-scale, expert-annotated retinal image datasets to detect referable or sight-threatening DR. However, publicly accessible documentation describing dataset provenance—such as population characteristics, imaging-device diversity, grading procedures, and handling of ungradable images—varies across systems. Reported architectural features occasionally include image-quality triage or lesion-aware modules, although descriptions differ in detail between manufacturers. Multimodal models or foundation-model–based approaches have not yet appeared among CE-certified autonomous DR screening systems and therefore fall outside the scope of this section.

Performance metrics publicly reported for CE-certified systems originate from individual validation studies commissioned or conducted by developers and independent groups. Consistent with a scoping-review methodology, we synthesized findings descriptively rather than pooling diagnostic accuracy estimates, due to heterogeneity in study design, reference standards, imaging hardware, and populations. RetCAD and Mona DR exhibit comparable performance in independent evaluations, with added flexibility for integration into national screening programs through modular deployment and real-time image quality feedback. Evidence specific to sight-threatening DR (STDR) is reported less consistently and shows wider variability, reflecting differences in study design and the constraints of single-modality fundus imaging.

Collectively, validation results suggest that CE-certified systems demonstrate diagnostic performance suitable for regulatory approval; however, heterogeneity across study methodologies precludes direct comparison. Documentation related to dataset characteristics, such as geographic representation, device heterogeneity, grading standards, class imbalance, and ungradable-image protocols, remains inconsistently reported across manufacturers. These contextual distinctions influence suitability for different screening scenarios, including primary care, teleophthalmology, national DR programs, and low-resource settings, and are therefore explicitly summarized in Table 2 and incorporated into the individual device descriptions.

ALGORITHM DESCRIPTIONS AND EVIDENCE ACCUMULATION

IDx-DR (Digital Diagnostics)

IDx-DR is an autonomous CNN-based system that employs a standardized inference pipeline composed of automated image-quality control, hierarchical lesion detection, and referable-DR classification. Consistent with other CNN-based systems

Table 1 CE-certified AI-based ophthalmic devices for diabetic retinopathy screening in the European Union

Device	Manufacturer	Primary function	MDR class	Autonomy	Representative evidence
IDx-DR	Digital Diagnostics	DR screening	IIa	Autonomous	Pivotal multicenter prospective studies
EyeArt	Eyenuk	DR / AMD / glaucoma detection	IIb	Autonomous	Multicenter European validation; JAMA Netw Open 2021
RetCAD	Thirona	DR + AMD dual detection	IIa	Assistive	European reader studies
Mona DR	MONA.health	DR detection with explainable output	IIa	Semi-autonomous	Independent EU evaluations
Remidio Medios AI	Remidio	Smartphone DR screening	II	Assistive	CE 2023; multicenter prospective validation
Retmarker DR	Retmarker	DR monitoring (microaneurysm turnover)	IIa	Assistive	EU screening projects and longitudinal evidence
SELENA +	EyRIS	DR + AMD + glaucoma triage	IIa	Semi-autonomous	CE 2020; real-world national program use
RetinoScan	TeleMedC	Autonomous DR screening	IIb	Autonomous	Cross-jurisdictional comparative studies
Aireen DR	Aireen Technologies	DR screening	IIb	Autonomous	Multicenter prospective trials in EU primary care
OphthAI	OphthAI Solutions (Evolucare / ADCIS)	DR + AMD + glaucoma detection	IIa	Autonomous	EU and Asian multicenter validations
LuxIA	LuxIA Diagnostics	DR screening	IIb	Autonomous	CE 2025; early deployment evaluations in EU primary-care screening

Table 1 continued

Device	Manufacturer	Primary function	MDR class	Autonomy	Representative evidence
Airdoc-Eye DR	Airdoc (CN)	Autonomous DR screening	IIb	Autonomous	Multicenter real-world deployment, including national screening networks
EyeWisdom (Vistel)	Visionary Intelligence (Vistel)	DR and other retinal abnormalities	IIa	Assistive	Embedded in multimodal fundus cameras; reader and workflow studies in EU sites

DR diabetic retinopathy, *AMD* age-related macular degeneration

described in this review, its design emphasizes robustness across heterogeneous camera inputs. non-mydriatic fundus photographs to detect and classify referable diabetic retinopathy (RDR). Its convolutional neural network (CNN) architecture localizes key retinal lesions—microaneurysms, hemorrhages, and exudates, through an end-to-end inference pipeline with automated image-quality control. After MDR conformity assessment, the device obtained CE marking (Class IIa) supported by multicenter prospective trials and post-market clinical follow-up. European studies have consistently shown sensitivity and specificity exceeding 90% for RDR versus masked expert graders, with robust performance across camera types and primary-care workflows. Field surveillance continues to monitor model drift, equity of performance, and safety alerts. Remaining challenges include limited publicly available information on dataset provenance, particularly geographic diversity and camera heterogeneity, and persistent variability in STDR sensitivity due to reliance on 2D fundus imaging. Enhanced transparency of change-control processes will be required to align with upcoming EU AI Act expectations.

EyeArt (Eyenuk)

EyeArt applies an ensemble of CNN classifiers with pre-processing and automated quality triage. Under the MDR, additional indications have been reported in regulatory summaries [10]. Multicenter European validation studies describe sensitivities and specificities for referable DR (RDR) within ranges generally consistent with regulatory expectations [6, 11]. Its post-market surveillance framework is documented within MDR-required PMS submissions [10] EyeArt has been evaluated in multiple geographic settings; however, additional transparency regarding training-data composition, demographic representativeness, and ungradable-image handling could further strengthen alignment with trustworthy-AI principles. Public reporting related to STDR definitions and evaluation methodologies remains limited in available documentation [17]. Further transparency regarding training-data composition, demographic representativeness, and ungradable-image handling would strengthen alignment with trustworthy-AI principles.

RetCAD (Thirona)

RetCAD uses multi-scale CNNs to jointly detect DR and AMD features from color fundus photographs and translate outputs into human-readable severity levels suitable for screening triage. Observational evidence from European reader-studies reports sensitivities and specificities within the ranges described in individual

Table 2 Extracted 13 CE-certified systems with key metrics

System	Company		Sensitivity (%)	Specificity (%)	Study type and context	Class	Certification year	Multi-disease
IDx-DR	Digital Diagnostics	90		90	Prospective multicenter trial; primary-care setting; multiple camera models	IIa	2022	No
EyeArt	Eyenuk	94		92	Prospective multicenter; US + EU cohorts; mixed camera types	IIb	2021	Yes
RetCAD	Thirona	86		94	Retrospective EU reader studies; mixed-device image sets	IIa	2022	Yes
Mona DR	MONA.health	95		90	Prospective EU multi-clinic validation; heterogeneous population	IIa	2023	No
Remidio Medios AI	Remidio	92		92	Prospective multicenter, including India + EU; smartphone-based imaging	II	2023	No
Retmarker DR	Retmarker	87		94	Real-world national program (Portugal); longitudinal data	IIa	2020	No
SELENA +	EyRIS	93		91	Multicenter Asia + Europe; real-world deployment	IIa	2020	Yes

Table 2 continued

System	Company		Sensitivity (%)	Specificity (%)	Study type and context	Class	Certification year	Multi-disease
RetinoScan	TeleMedC	91		90	Prospective Australia + EU cohorts; cross-camera compatibility	IIb	2022	No
Aireen DR	Aireen Technologies	92		91	Prospective EU primary-care multicenter trial	IIb	2024	No
OphthAI	Evolucare/ADCIS	93		92	Prospective EU + Asia cohorts; multi-camera validation	IIa	2023	Yes
LuxIA	LuxIA Diagnostics	90		93	Technical performance + early EU prospective deployment	IIb	2025	No
Airdoc-Eye DR	Airdoc	91.8		93.1	Large multi-ethnic prospective cohorts; community + real-world settings	IIb	2024	Yes
EyeWisdom (Vistel)	Visionary Intelligence	89–94		88–93	Reader studies; embedded in multimodal fundus-camera platforms; EU deployments	IIa	2021	Yes

publications. The system is CE-certified as Class IIa under MDR after clinical evaluation confirming conformity with safety and performance criteria. Strengths and limitations are variably documented, and publicly accessible information on dataset diversity, annotation procedures, and STDR performance remains limited.

Mona DR (MONA.health)

Mona DR combines pixel-level CNN lesion extraction with contextual image analysis based on vascular and topological features. Independent European assessments report performance metrics for RDR consistent with

the methodologies and case definitions used within each study [11]. The device is CE-marked and subject to ongoing post-market performance monitoring under MDR. Advantages and remaining evidence gaps relate to publicly accessible reporting on dataset composition, image-source diversity, and subgroup analyses.

Remidio Medios AI (Remidio)

Remidio Medios AI is a smartphone-based deep learning system designed for automated detection of RDR from non-mydriatic fundus images. The platform integrates a portable fundus camera with embedded on-device inference, enabling offline operation in primary-care and community settings without reliance on high-bandwidth Internet. The algorithm employs convolutional neural networks (CNNs) trained on multi-ethnic image datasets, coupled with an integrated image-quality assessment module to exclude ungradable images. Following clinical validation across Indian and European populations, the system received CE marking under the EU MDR (Class II) in 2023. Reported sensitivity and specificity for RDR detection are drawn from prospective and retrospective studies and fall within ranges reported in published evaluations. Its portability, low-cost implementation, and absence of cloud dependency are described in technical reports; however, long-term post-market evidence is still emerging, and transparency regarding dataset composition and STDR-specific performance remains limited, particularly as EU AI Act requirements for dataset governance expand. Remidio Medios AI is also authorized in India, although this falls outside the scope of EU regulatory assessment.

Retmarker DR (Retmarker)

Retmarker DR is a CE-certified (Class IIa) software system that supports diabetic retinopathy screening and longitudinal disease-progression monitoring through analysis of microaneurysm turnover and other fundus-based biomarkers. Unlike purely image-classification AIs, Retmarker emphasizes temporal tracking by quantifying the appearance and disappearance of microaneurysms between visits, providing surrogate indicators of retinopathy activity. Its architecture combines classical image-processing pipelines with deep-learning lesion detection to improve robustness in variable image conditions. The system has been deployed in several EU public-health screening programs, including the Portuguese National DR Screening Network, and validated in observational studies reporting performance values as described within those studies for RDR identification compared with expert graders. Post-market evaluations confirm strong reproducibility and reliability across imaging devices. Strengths and evidence gaps relate to publicly available information on dataset provenance, device diversity, and STDR-specific performance.

SELENA + (EyRIS)

SELENA + is a deep-learning platform capable of detecting referable DR, age-related macular degeneration (AMD), and glaucoma-suspect findings from color fundus photographs. It uses an ensemble of CNN classifiers with hierarchical decision layers that output both disease probability scores and visual heat maps for explainability. The system achieved CE marking in 2020 under the MDR (Class IIa) and represents one of the first multi-disease ophthalmic AI solutions recognized in Europe. Multicenter validations in Asia and Europe report performance metrics within ranges published in individual studies for RDR detection relative to ophthalmologist consensus. Real-world deployments within national screening programs, including the Singapore National DR Screening Program, have been described in national screening programs in primary-care workflows. Its key advantages and limitations relate to public availability of dataset provenance, glaucoma-suspect evaluation details, and harmonized post-market reporting, which will become increasingly relevant under the EU AI Act.

RetinoScan (TeleMedC)

RetinoScan is an autonomous AI system designed for DR screening using non-mydriatic fundus cameras. The model employs

convolutional networks with integrated attention modules to focus on diagnostically salient retinal regions and includes an automated image-quality triage step. The device received CE certification (Class IIb) under the MDR and is concurrently cleared by the Australian TGA (Class IIa), illustrating cross-jurisdictional regulatory recognition. In clinical evaluations across Australia and Europe, RetinoScan reports performance values consistent with study methodologies and reference standards. Workflow integration studies demonstrate efficient point-of-care deployment in primary clinics with minimal operator training. Strengths include full autonomy, compatibility with multiple camera types, and cross-region regulatory compliance; remaining challenges include limited published evidence on STDR performance, dataset representativeness, and long-term post-market surveillance outcomes.

Aireen DR (Aireen Technologies)

Aireen DR is an autonomous deep learning system designed to detect RDR in non-mydriatic fundus photographs [18]. It leverages a CNN with multi-layer feature extraction and an integrated image-quality assessment module to ensure reliable input for inference. The system achieved CE marking as a Class IIb device under the EU MDR in 2024, supported by prospective multicenter trials across European primary-care settings. Validation studies present performance estimates within ranges documented in the corresponding publications Aireen DR's post-market surveillance framework, compliant with MDR requirements, includes periodic safety update reports (PSURs) to monitor model drift and ensure equitable performance across ethnic and demographic subgroups. Key strengths include its user-friendly interface and compatibility with low-resource settings, though remaining limitations are described in technical documentation, while remaining evidence gaps include STDR evaluation in low-prevalence contexts and transparency regarding dataset provenance, which are relevant for forthcoming EU AI Act requirements.

OphthAI (OphthAI Solutions)

OphthAI is a multi-disease AI platform that detects referable DR, AMD, and glaucoma-suspect features from color fundus images using an ensemble of CNNs augmented with attention mechanisms for enhanced lesion localization [19]. The system provides explainable outputs through visual heatmaps, improving clinical interpretability. CE-marked as a Class IIa device under the MDR in 2023, OphthAI underwent rigorous clinical evaluation in European and Asian cohorts, reporting performance values consistent with the design and populations of those studies. Its MDR-compliant post-market surveillance includes continuous performance monitoring and PSUR submissions to track algorithm reliability across varied imaging conditions. OphthAI's strengths lie in its multi-disease detection capabilities and seamless integration into teleophthalmology workflows. However, improvements in STDR threshold harmonization, dataset-representativeness reporting, and CONSORT-AI-aligned transparency will be essential as EU AI Act requirements become operational.

LuxIA (LuxIA Diagnostics)

LuxIA is a newly CE-certified (Class IIb) AI-based system developed by LuxIA Diagnostics for autonomous detection of referable diabetic retinopathy (RDR) and quantitative lesion analysis. The platform integrates convolutional neural networks (CNNs) with transformer-based feature extraction to enhance recognition of subtle microaneurysms and exudates across varying image qualities. In addition to binary RDR classification, LuxIA provides lesion quantification metrics—such as microaneurysm count and hemorrhage area—to support disease-severity grading and longitudinal monitoring.

Following conformity assessment under the EU Medical Device Regulation (MDR), LuxIA received CE marking in 2025 after independent European technical-performance evaluations report performance metrics as described in those studies. The system's semi-autonomous

workflow allows integration into existing DR screening pipelines, providing visual heatmaps for interpretability and automated image-quality control. Ongoing post-market surveillance focuses on real-world robustness, interoperability across diverse fundus-camera models, and harmonization of reporting standards with the forthcoming EU AI Act. As with other systems, publicly available information on dataset composition and STDR performance remains limited.

Airdoc-Eye DR (Airdoc, China)

Airdoc-Eye DR is an autonomous deep learning system for detectiing referable diabetic retinopathy (RDR) in non-mydriatic fundus photographs. The platform employs an ensemble of convolutional neural networks with lesion-attention mechanisms to identify microaneurysms, hemorrhages, and exudates, alongside an integrated image-quality triage module. Airdoc is notable for being the first ophthalmic AI screening system to obtain Class III approval from the National Medical Products Administration (NMPA) in China, reflecting its authorization for autonomous clinical decision-making without mandatory ophthalmologist confirmation. The system subsequently achieved CE certification under the EU MDR as a Class IIb medical device, enabling deployment within European DR screening workflows.

Multicenter prospective studies in China and Southeast Asia report performance metrics consistent with their respective methodologies, with stable performance across community, primary-care, and teleophthalmology environments. Real-world deployment has occurred at scale, including integration into provincial DR screening programs and primary-care diabetes management pathways. Airdoc's strengths include regulatory maturity in autonomous diagnosis, large-scale multiethnic validation cohorts, and demonstrated operational scalability. Ongoing challenges and limitations relate to public reporting of dataset composition, STDR-subgroup analyses, and change-control documentation, which will be relevant under the EU AI Act.

Vistel/EyeWisdom (Visionary Intelligence Ltd.)

EyeWisdom is a cloud-connected AI platform for retinal screening developed by Visionary Intelligence (Vistel), originally deployed in China and later integrated into European multimodal screening devices such as the Visionix VX610 and VX650. The system uses deep-learning classifiers with built-in image-quality assessment to detect multiple posterior-segment pathologies, including diabetic retinopathy, exudative and non-exudative AMD, glaucoma-suspect changes, and "other abnormalities", from color fundus photographs, generating structured PDF reports that can be incorporated into primary-care and optical-retail workflows [20]. In European deployments, EyeWisdom is marketed as AI "medical-device software" embedded in retinal cameras, supporting opportunistic case finding rather than fully autonomous diagnosis, and thus typically positioned as a triage or decision-support tool with ophthalmologist over-read. Published evaluations from Asian and European settings report high diagnostic accuracy for referable DR and other retinal disease categories, though performance estimates vary across study designs and reference standards; real-world studies have also explored the cost-effectiveness of EyeWisdom-based screening in low-resource settings. Remaining gaps include limited publicly accessible information on training-data provenance, demographic composition, camera diversity, and sight-threatening DR (STDR)–specific performance, as well as sparse CONSORT-AI–style reporting of prospective implementations. As EyeWisdom and related Vistel products expand into EU markets, clearer documentation of dataset governance, change-control procedures, and post-market surveillance will be important to demonstrate alignment with MDR obligations and forthcoming EU AI Act transparency and fairness requirements.

DISCUSSION

Overview of AI Architectures Used in CE-Certified DR Screening Systems

To provide a unified technical foundation for interpreting the differences among CE-certified diabetic retinopathy (DR) AI systems, we summarize the principal architectural families reported in CE-certified and related DR-screening literature. Most DR screening devices deploy one of four core architectures, including convolutional neural networks (CNNs), vision transformers (ViTs), ensemble or hybrid models, and attention-based mechanisms.

Convolutional Neural Networks (CNNs) remain the dominant architecture in CE-certified DR systems. CNNs extract hierarchical spatial features (e.g., microaneurysms, hemorrhages, exudates) using progressively deeper convolutional layers, and have been widely applied to fundus-image classification tasks. Their established use in ophthalmic imaging may contribute to their continued prominence among currently approved systems.

Vision Transformers (ViTs) and transformer–CNN hybrids represent a newer generation of architectures capable of modeling long-range global context and subtle texture patterns. Although fewer CE-marked devices currently use ViTs, early evaluations outside the CE-certified landscape have explored their potential for capturing global retinal context and subtle image patterns, although evidence in regulated DR-screening settings remains limited.

Ensemble and Hybrid Models combine multiple CNNs or integrate multimodal components (e.g., quality-triage networks, lesion detectors, or disease-specific classifiers). Such configurations have been reported to support stability across heterogeneous image inputs, although the extent to which ensembling contributes to real-world performance is not uniformly documented in CE-certified devices.

Attention Mechanisms, either as standalone modules or integrated within CNNs/Transformers, focus the model on clinically salient retinal regions. These mechanisms are often used to highlight image regions deemed important by the model and may support visual interpretability. However, the specific role of attention modules in meeting emerging transparency or explainability expectations is infrequently detailed in public regulatory documentation.

Although not MDR-certified, these systems provide important historical context for how automated DR grading evolved. Understanding these architectural families provides context for interpreting the heterogeneous design choices observed across CE-certified systems. Table 3 provides a structured comparison of the principal architectures and their clinical implications. Beyond currently CE-certified systems, emerging work on vision foundation models and multimodal fundus–OCT architectures has been reported in research settings; these technologies are noted here only to contextualize the broader technical landscape and are not the focus of this scoping review.

Dataset Provenance, Diversity, and Transparency Considerations

A key determinant of real-world performance in AI-based diabetic retinopathy screening is the provenance and diversity of the datasets used for model development and validation. Across CE-certified systems, publicly available information demonstrates considerable heterogeneity in dataset characteristics. Most systems were trained on large-scale retrospective image repositories, yet the documented degree of geographic and ethnic diversity varies widely, with some models validated primarily in single-region European cohorts while others incorporate multi-ethnic datasets from Asia, North America, or mixed global registries. Camera variability, including different sensor types, fields of view, and image-quality distributions, is inconsistently reported, despite its known impact on model robustness and generalizability.

Ground-truth adjudication processes also differ among devices. While several systems describe multi-grader consensus or adjudication by retinal specialists, others provide limited detail on grader expertise, disagreement resolution, or disease-severity scales used. In addition, class imbalance, particularly the relative scarcity

Table 3 Summary of major AI architecture classes used in CE-certified DR screening systems and their clinical implications

Architecture class	Core computational features	Strengths	Limitations	Clinical implications
Convolutional Neural Networks (CNNs)	Hierarchical local feature extraction; convolutional filters detect microaneurysms, hemorrhages, exudates	High accuracy; stable across cameras; computationally efficient; widely validated	Less effective at modeling long-range global structures; sensitivity may drop in low-contrast cases	Reliable for large-scale DR screening; strong real-world evidence base; widely used in CE devices
Vision Transformers (ViTs)	Global self-attention; patch-based global context modeling	Superior modeling of diffuse changes; potentially higher generalizability; strong scalability	Requires larger datasets; fewer CE-validated devices; more computationally intensive	Promising for next-generation DR screening, especially multimodal workflows
Ensemble / Hybrid Models	Combines multiple backbones (CNN + lesion detector + quality triage); weighted decision fusion	Improved robustness; mitigates single-model errors; enhances stability in primary-care workflows	More complex to validate; opaque architecture to regulators	Useful for MDR compliance due to higher stability and reduced variance in performance
Attention-Enhanced Models	Modules highlight structurally important retinal regions	Better explainability; focuses on clinically relevant lesions; reduces noise impact	Attention maps may not always reflect pathology accurately	Supports EU AI Act requirements on transparency and human interpretability
Multimodal Models (Emerging)	Combines fundus images + OCT + text/clinical data	Comprehensive disease profiling; improved STDR detection	Not yet widely CE-certified; high data requirements	Expected direction for future regulatory submissions and clinical translation

of sight-threatening DR (STDR), is rarely quantified, despite being a key driver of model sensitivity in real-world settings. The handling of ungradable images, an important element of MDR-compliant workflow design, is also variably described, even though it can influence referral accuracy and workload.

Dataset provenance, including population composition, geographic distribution, camera heterogeneity, annotation processes, and class imbalance, varies substantially across CE-certified systems. Many manufacturers provide only high-level summaries of training datasets, with limited disclosure of demographic representation, adjudication protocols, or ungradable-image handling. These gaps hinder rigorous assessment of model fairness, robustness, and generalizability, and present challenges for MDR-mandated PSUR reporting and the EU AI Act's requirements for transparent and representative datasets. For legacy and non-MDR systems such as EyeWisdom (Vistel), publicly available information on dataset composition is even more limited, underscoring why they were treated descriptively rather than as part of the main CE-certified cohort. By synthesizing what is known and clearly identifying what remains unknown, this review highlights critical areas where improved reporting standards, such as CONSORT-AI and SPIRIT-AI, would substantially enhance the reliability and accountability of ophthalmic AI systems.

Practical Deployment Requirements for CE-Certified DR AI Systems

Successful implementation of CE-certified AI-based DR screening tools depends on multiple factors beyond initial regulatory approval. Real-world deployment depends on a series of operational, technical, and governance prerequisites that ensure consistent performance across diverse clinical environments. First, camera compatibility must be verified, including supported sensor types, fields of view, and minimum image-quality thresholds. Many systems require standardized illumination and focus parameters, making image-quality triage protocols essential to reduce false positives and avoid ungradable outputs. Integration into routine care also relies on interoperability with EHR/PACS systems, secure data transfer workflows, and the ability to embed the AI decision into existing screening pathways.

Equally important is operator training, which includes correct image acquisition techniques, device troubleshooting, and understanding of referral recommendations. MDR-classified systems must also comply with cybersecurity and data-protection requirements, ensuring encrypted transmission, audit logs, and protection against adversarial interference. Finally, real-world performance must be continuously monitored through post-market surveillance (PMS) mechanisms mandated by the MDR, such as periodic safety update reports (PSURs), vigilance reporting, and performance audits across subpopulations and camera types. Together, these elements illustrate that AI deployment operates at a systems level, involving interaction between clinicians, program administrators, and manufacturers.

Figure 1 offers a schematic illustration of how evidence for AI-based screening systems may progress from early technical development through clinical studies toward post-market evaluation and lifecycle monitoring within MDR and EU AI Act frameworks. Findings from the included sources suggest that multiple CE-certified AI systems have been evaluated in a variety of clinical and screening settings across Europe and other regions, with differing study designs, endpoints, and reporting conventions [21]. However, publicly available information on training-data provenance, such as population diversity, camera variability, and grading standards, remains inconsistent across manufacturers, limiting full assessment of generalizability [22, 23]. Nonetheless, STDR detection remains a relative weakness across devices, emphasizing the need for further algorithmic refinement, broader dataset curation with explicit documentation of representativeness, prospective head-to-head trials at matched operating points, and inclusion of multimodal data (e.g., OCT) where feasible [24, 25]. These limitations underline that current fundus-only CE-certified systems share structural constraints, rather than vendor-specific deficiencies.

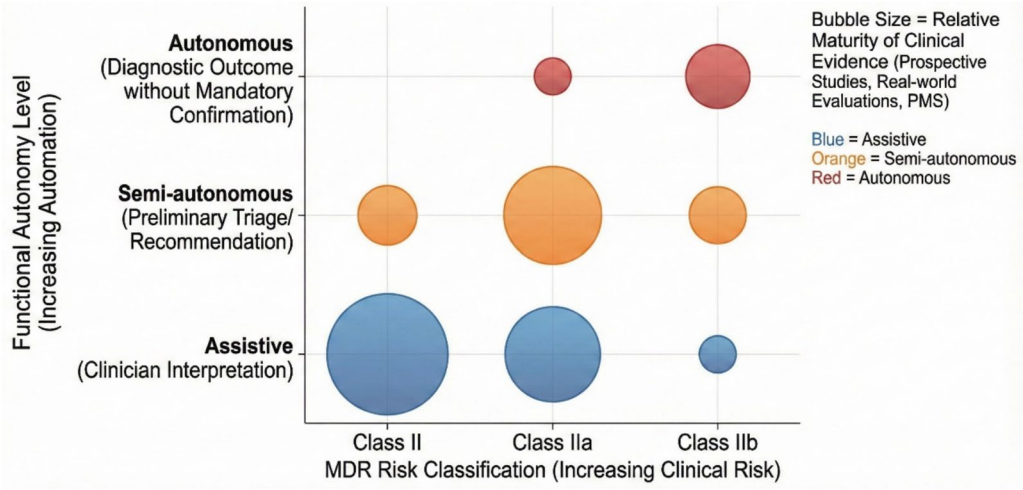

Fig. 1 Outline of the progressive evidence pathway for AI-based DR screening systems, moving from technical development and retrospective validation to prospective clinical studies and real-world deployment. Evidence requirements increase across stages, emphasizing data quality, transparency, and ongoing performance monitoring under the MDR and EU AI Act. *PMS* post-market surveillance

Regulatory trajectories under the MDR and forthcoming AI Act will place greater emphasis on transparency, logging, human oversight, and risk management for high-risk AI. Adoption of CONSORT-AI/SPIRIT-AI reporting, trustworthy-AI principles, and good machine-learning practice is likely to improve reproducibility, reduce bias, and build trust among clinicians and patients. Evidence packages should include subgroup analyses and program-level metrics (e.g., referral yield, time-to-treatment), and clearly documented model-change governance to ensure safe evolution over time [26]. In particular, the AI Act's requirements for traceable dataset documentation, continuous monitoring of real-world performance, and structured change-control reporting are likely to inform future regulatory submissions by ophthalmic AI manufacturers [14, 27].

Recent comparative studies provide further context for understanding algorithmic variability. In a Polish multicenter cohort of 750 participants, both IDx-DR and Mona DR achieved 100% sensitivity for STDR when real-time image-retake functionality was enabled, confirming the robustness of autonomous workflows under controlled acquisition conditions [28]. Similarly, a head-to-head comparison between IDx-DR and RetCAD revealed high agreement for referable DR but demonstrated threshold-dependent trade-offs: IDx-DR favored higher sensitivity, while RetCAD achieved superior specificity, illustrating the importance of calibration and reference-standard alignment [29]. In contrast, large-scale real-world evaluations show wider inter-system variability.

The Veterans Affairs multicenter study involving over 23,000 patients reported sensitivities ranging from 50.9 to 85.9% across seven commercial algorithms [30], while the Finnish Study found mean sensitivity of 77.5%, specificity of 80.6%, and an ungradable-image rate of 1.9% among 21 CE-certified AI models [31]. Consistent results were also seen in Jin et al., 2021, where direct comparison of IDx-DR and RetinaLyze demonstrated 93–95% agreement, confirming high diagnostic concordance yet highlighting modest differences in decision thresholds [32]. Together, these findings underscore that performance discrepancies among CE-certified systems are driven largely by differences in dataset composition, imaging hardware, annotation standards, and threshold calibration rather than intrinsic algorithmic superiority [33]. This further supports the need for harmonized evaluation frameworks

and transparent reporting aligned with MDR and the future AI Act. While the present review does not conduct comparative performance ranking, inclusion of Airdoc-Eye DR within the landscape illustrates the increasing globalization of ophthalmic AI and the heterogeneity of regulatory pathways.

Incorporating Airdoc-Eye DR into the comparative landscape further highlights the increasing globalization and regulatory diversification of autonomous ophthalmic AI. Unlike most CE-certified systems that have primarily advanced through the EU MDR pathway, Airdoc-Eye DR is distinguished by its China NMPA Class III approval, authorizing fully autonomous referral decisions without mandatory ophthalmologist oversight. This designation reflects a comparatively higher regulatory confidence threshold and demonstrates that autonomous diagnostic workflows can reach national-scale deployment when post-market surveillance (PMS), system auditability, and operator training frameworks are robust.

Given the rapid expansion of AI-enabled DR screening across global regions, understanding regulatory diversity is essential for interpreting system maturity and deployment readiness. Figure 2 presents a comparative overview of regulatory pathways in China, the European Union, and Australia. China's NMPA allows fully autonomous diagnostic AI systems under Class III designation, while the EU MDR regulates most DR-AI devices as Class IIa or IIb, requiring Notified Body assessment, clinical evaluation, and post-market surveillance. Australia's TGA applies a Class IIa SaMD framework. These distinctions demonstrate how differing regulatory philosophies shape approval thresholds, evidence expectations, and real-world implementation strategies for ophthalmic AI technologies.

In Europe, Airdoc-Eye DR has also obtained CE Class IIb certification, placing it in the same high-risk clinical use category as systems such as EyeArt, RetinoScan, and OphthAI. Comparative evaluations suggest performance levels comparable to IDx-DR, Mona DR, and Aireen

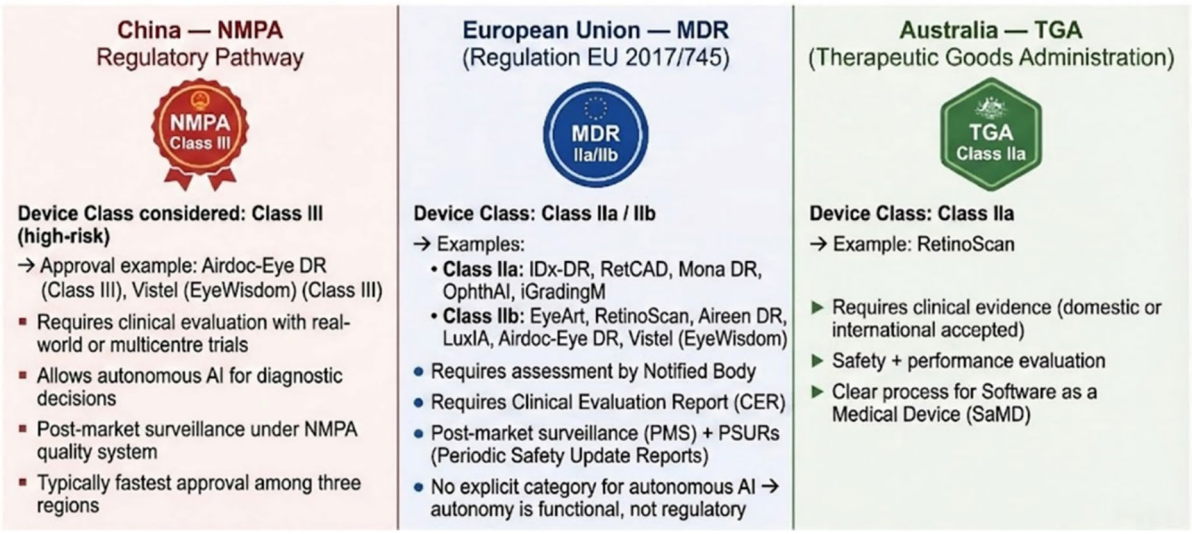

Fig. 2 Comparative overview of global regulatory pathways governing AI-based diabetic retinopathy screening devices in China, the European Union, and Australia. The schematic contrasts the classification systems and approval requirements across major jurisdictions. China's NMPA permits autonomous diagnostic AI systems under Class III medical device designation, exemplified by Airdoc-Eye DR and Vistel (EyeWisdom). In the EU, AI-based ophthalmic systems are classified under MDR as Class IIa or IIb devices, requiring Notified Body assessment, clinical evaluation, and post-market surveillance. Australia's TGA regulates these systems under Class IIa SaMD requirements. This comparison illustrates the convergent but distinct regulatory philosophies shaping the deployment of ophthalmic AI worldwide

DR for detecting referable diabetic retinopathy (DR), with > 90% sensitivity and specificity demonstrated across controlled trial settings and community-based screening programs. However, consistent with other 2D fundus-based autonomous systems, reduced sensitivity for sight-threatening DR (STDR) persists in the absence of OCT or multimodal imaging, emphasizing that STDR detection limitations are structural to single-modality analysis rather than vendor-specific. This reinforces the broader field-wide need for multimodal pipelines, including fundus–OCT fusion and foundation-model architectures currently emerging in research settings [34, 35].

Notably, China is rapidly expanding domestic authorization of AI-based ophthalmic diagnostic tools, illustrating a parallel but distinct regulatory trajectory. Recent approvals include EyeWisdom (Vistel) and Airdoc-AI FUNDUS, which have progressed through China's accelerated clinical evaluation pathways for AI-enhanced medical devices [36]. This trend suggests an emerging ecosystem in which multi-disease, multi-region, and real-world scalable DR screening platforms are becoming feasible across global health systems, shaped by divergent regulatory philosophies but convergent goals of improving access, efficiency, and care standardization. Such developments highlight the urgency for EU regulators and developers to enhance dataset transparency and harmonized reporting to ensure safe cross-border deployment [5, 37].

Despite these advances, most ophthalmic AI systems considered in this review, whether CE-certified in the EU or authorized under other jurisdictions, have been developed primarily for DR-related tasks using 2D fundus images. This task focus influences the range of conditions that can be addressed and shapes expectations for generalizability across populations, devices, and care settings. Continuous model governance—including drift monitoring, documentation of software updates, and transparency around algorithm changes—is increasingly highlighted within MDR and EU AI Act discussions. The MDR and forthcoming AI Act place increasing emphasis on dataset documentation, representativeness, and traceability. Transparent reporting of population composition, camera diversity, annotation methods, and ungradable-image handling will be essential for future regulatory submissions. Future development directions proposed in the literature include multi-task and multi-disease frameworks, multimodal inputs, and uncertainty quantification, alongside adherence to reporting guidance such as CONSORT-AI and SPIRIT-AI [38, 39]. These measures will be crucial to enhance reproducibility, minimize bias, and strengthen clinician and patient trust in autonomous ophthalmic AI systems.

Current CE-certified systems constitute the first wave of DR-screening AI devices to receive CE certification. As technologies evolve, expectations regarding transparency, model governance, and post-market evidence generation under the MDR and AI Act are also likely to evolve.

Emerging Trends and Future Directions in AI-Based DR Screening

Beyond currently CE-certified systems, several emerging technological pathways are expected to shape the next generation of diabetic retinopathy (DR) screening tools. First, OCT-based and OCT-integrated AI models have been proposed to improve the detection of sight-threatening DR and diabetic macular edema—limitations inherent to 2D fundus photography. Early clinical evaluations suggest that multimodal fundus + OCT architectures can enhance sensitivity for macular pathology and offer more granular disease-severity stratification. Second, progression-prediction algorithms, including models that track longitudinal structural change or estimate individualized response to therapy, have been explored in personalized DR management. Third, multi-disease generalist platforms capable of detecting DR, age-related macular degeneration, glaucoma-suspect features, and systemic-disease markers from retinal images are expected to expand, aligning with broader population-screening strategies. How such systems will be evaluated within MDR and EU AI Act frameworks remains an area of active discussion.

Several of these technologies are being discussed in the context of future MDR and AI Act regulatory pathways, with particular attention

to data governance, transparency, explainability, and robust post-market surveillance. As regulatory frameworks evolve, AI systems incorporating uncertainty quantification, continual learning safeguards, and multimodal interoperability have been proposed as potential candidates for future clinical implementation. Taken together, these emerging directions suggest a gradual shift from single-task image classifiers toward more comprehensive and adaptively governed platforms, with corresponding implications for dataset governance, continual-learning controls, and transparency documentation. Future regulatory submissions are likely to emphasize dataset governance, continual-learning controls, transparency documentation, uncertainty quantification, and model retraining safeguards, requirements that are only beginning to be defined in the current regulatory landscape.

Several limitations of this scoping review should be acknowledged. First, given the descriptive and regulatory focus of this work, the synthesis is necessarily based on heterogeneous sources, including regulatory filings, manufacturer documentation, and peer-reviewed studies with variable levels of methodological detail. As a result, direct head-to-head comparison of performance metrics across CE-certified DR screening systems is not appropriate. Reported sensitivity, specificity, and ungradable-image rates are derived from studies using different datasets, population compositions, grading reference standards, disease definitions, and operating thresholds, which preclude meaningful comparative ranking or inference of algorithmic superiority. Second, publicly available information on training-data provenance, demographic representation, camera diversity, and adjudication protocols remains incomplete for many systems, limiting formal assessment of fairness, bias, and generalizability. Third, this review does not perform a meta-analysis or pooled performance estimation, as study heterogeneity and inconsistent reporting would violate key assumptions required for quantitative synthesis. Finally, rapidly evolving regulatory requirements under the MDR and forthcoming EU AI Act mean that some systems discussed here may undergo post-certification modifications or reclassification, and reported characteristics may change over time. These limitations underscore the need for harmonized evaluation frameworks, transparent dataset documentation, and prospective studies conducted at matched operating points to enable robust comparative assessment in future work.

CONCLUSIONS

CE-certified AI-based ophthalmic devices now constitute an established, though still evolving, component of DR screening workflows in parts of Europe. The presented systems demonstrate strong diagnostic performance and have obtained CE certification under the MDR. However, dataset transparency, multimodal integration, and harmonization of performance reporting remain essential areas for advancement. Continued progress is likely to require rigorous post-market evaluation, improvements in reporting standards, and collaboration among clinicians, regulators, and developers to enhance transparency, accountability, and interpretability of AI-assisted ophthalmic screening.

Author Contributions. Both authors contributed substantially to the conception and design of the work. Drafting of the manuscript: Andrzej Grzybowski and Kai Jin. Critical revision of the manuscript for important intellectual content: Andrzej Grzybowski and Kai Jin. Final approval of the version to be published: Both authors.

Funding. No funding or sponsorship was received for the publication of this article.

Data Availability. Data sharing is not applicable to this article, as no datasets were generated or analyzed during the current study.

Declarations

Conflict of Interest. Andrzej Grzybowski is an Editorial Board member of Ophthalmology and Therapy and was not involved in the selection of peer reviewers for this manuscript nor

in any subsequent editorial decisions. Kai Jin declares no conflicts of interest.

Ethical Approval. This article is based on previously conducted studies and does not contain any new studies with human participants or animals performed by either author.

REFERENCES

1. Jin K, Li Y, Wu H, Tham YC, Koh V, Zhao Y, et al. Integration of smartphone technology and artificial intelligence for advanced ophthalmic care: a systematic review. Adv Ophthalmol Pract Res. 2024;4(3):120–7.

2. Grzybowski A, Singhanetr P, Nanegrungsunk O, Ruamviboonsuk P. Artificial intelligence for diabetic retinopathy screening using color retinal photographs: from development to deployment. Ophthalmol Ther. 2023;12(3):1419–37.

3. Hill L, Makaroff LE. Early detection and timely treatment can prevent or delay diabetic retinopathy. Diabetes Res Clin Pract. 2016;120:241–3.

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

5. González-Gonzalo C, Thee EF, Klaver CCW, Lee AY, Schlingemann RO, Tufail A, et al. Trustworthy AI: closing the gap between development and integration of AI systems in ophthalmic practice. Prog Retin Eye Res. 2022;90:101034.

6. Fraser AG, Biasin E, Bijnens B, Bruining N, Caiani EG, Cobbaert K, et al. Artificial intelligence in medical device software and high-risk medical devices—a review of definitions, expert recommendations and regulatory initiatives. Expert Rev Med Devices. 2023;20(6):467–91.

7. Beckers R, Kwade Z, Zanca F. The EU Medical Device Regulation: implications for artificial intelligence-based medical device software in medical physics. Phys Med. 2021;83:1–8.

8. Evans NG, Wenner DM, Cohen IG, Purves D, Chiang MF, Ting DSW, et al. Emerging ethical considerations for the use of artificial intelligence in ophthalmology. Ophthalmol Sci. 2022;2(2):100141.

9. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digit Med. 2018;1:39.

10. Fehr J, Citro B, Malpani R, Lippert C, Madai VI. A trustworthy AI reality-check: the lack of transparency of artificial intelligence products in healthcare. Front Digit Health. 2024;6:1267290.

11. Keutzer L, Simonsson US. Medical device apps: an introduction to regulatory affairs for developers. JMIR mHealth Uhealth. 2020;8(6):e17567.

12. Le H, Sindahl P, Andersen M, Hallgreen CE. Understanding the work of the Pharmacovigilance Risk Assessment Committee (PRAC): a quantitative review of the post-authorisation safety evaluation of antidiabetic drugs from 2012 to 2022. Drug Saf. 2025;48(7):781–94.

13. Jawahar N, Ashok NA, Nidamanuri BSS, Venkatachalam S, Selvaraj J. A review of the new European medical device regulations. Pharm Regul Aff. 2022;11(3):298.

14. Gasser U. An EU landmark for AI governance. Science. 2023;380(6651):1203.

15. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Lancet Digit Health. 2020;2(10):e537–48.

16. Rivera SC, Liu X, Chan AW, Denniston AK, Calvert MJ. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. BMJ. 2020;370:m3210.

17. Ong AY, Hogg HDJ, Kale AU, Taribagil P, Kras A, Dow E, et al. AI as a medical device for ophthalmic imaging in Europe, Australia, and the United States: protocol for a systematic scoping review of regulated devices. JMIR Res Protoc. 2024;13:e52602.

18. Šín M, Ženíšková R, Slíva M, Dvořák K, Vaľková J, Bayer J, et al. Comparison of the Aireen system with telemedicine evaluation by an ophthalmologist—a real-world study. Clin Ophthalmol. 2025. https://doi.org/10.2147/OPTH.S511233.

19. Roubelat FP, Soler V, Varenne F, Gualino V. Real-world artificial intelligence-based interpretation of fundus imaging as part of an eyewear prescription renewal protocol. J Fr Ophtalmol. 2024;47(6):104130.

20. Wang Y, Yu M, Hu B, Jin X, Li Y, Zhang X, et al. Deep learning-based detection and stage grading for optimising diagnosis of diabetic retinopathy. Diabetes Metab Res Rev. 2021;37(4):e3445.

21. Wang TW, Luo WT, Tu YK, Chou YB, Wu YT. Diagnostic accuracy of EyeArt for fundus-based detection of diabetic retinopathy: a systematic review and meta-analysis. Am J Ophthalmol. 2026;281:465–79.

22. Mahesh N, Devishamani CS, Raghu K, Mahalingam M, Bysani P, Chakravarthy AV, et al. Advancing healthcare: the role and impact of AI and foundation models. Am J Transl Res. 2024;16(6):2166–79.

23. Abràmoff MD, Tarver ME, Loyo-Berrios N, Trujillo S, Char D, Obermeyer Z, et al. Considerations for addressing bias in artificial intelligence for health equity. NPJ Digit Med. 2023;6(1):170.

24. Jin K, Yan Y, Chen M, Wang J, Pan X, Liu X, et al. Multimodal deep learning with feature-level fusion for identification of choroidal neovascularization activity in age-related macular degeneration. Acta Ophthalmol. 2022;100(2):e512–20.

25. Kang EY-C, Yeung L, Lee Y-L, Wu C-H, Peng S-Y, Chen Y-P, et al. A multimodal imaging-based deep learning model for detecting treatment-requiring retinal vascular diseases: model development and validation study. JMIR Med Inform. 2021;9(5):e28868.

26. Macdonald T, Zhelev Z, Liu X, Hyde C, Fajtl J, Egan C, et al. Generating evidence to support the role of AI in diabetic eye screening: considerations from the UK National Screening Committee. Lancet Digit Health. 2025;7(5):100840.

27. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. Lancet. 2019;393(10181):1577–9.

28. Grzybowski A, Peeters F, Barão RC. Evaluating the efficacy of AI systems in diabetic retinopathy detection: a comparative analysis of Mona DR and IDx-DR. Acta Ophthalmol. 2025;103(4):388–95.

29. Grzybowski A, Brona P, Krzywicki T, Ruamviboonsuk P. Diagnostic accuracy of automated diabetic retinopathy image assessment software: IDx-DR and RetCAD. Ophthalmol Ther. 2025;14(1):73–84.

30. Lee AY, Yanagihara RT, Lee CS, Blazes M, Jung HC, Chee YE, et al. Multicenter, head-to-head, real-world validation study of seven automated artificial intelligence diabetic retinopathy screening systems. Ophthalmology. 2021;44(5):1168–75.

31. Heloterä H, Arffman M, Sund R, Keskimäki I, Kaarniranta K. The incidence and prevalence of diabetic macular edema and proliferative diabetic retinopathy, their progression to visual impairment and patterns in their intravitreal treatment in the Finnish population. Acta Ophthalmol. 2024;102(6):643–52.

32. Grzybowski A, Brona P. Analysis and comparison of two artificial intelligence diabetic retinopathy screening algorithms in a pilot study: IDx-DR and Retinalyze. J Clin Med. 2021. https://doi.org/10.3390/jcm10112352.

33. Martinez-Martin N, Luo Z, Kaushal A, Adeli E, Haque A, Kelly SS, et al. Ethical issues in using ambient intelligence in health-care settings. Lancet Digit Health. 2021;3(2):e115–23.

34. Zhou Y, Chia MA, Wagner SK, Ayhan MS, Williamson DJ, Struyven RR, et al. A foundation model for generalizable disease detection from retinal images. Nature. 2023;622(7981):156–63.

35. Qiu J, Wu J, Wei H, Shi P, Zhang M, Sun Y, et al. Development and validation of a multimodal multitask vision foundation model for generalist ophthalmic artificial intelligence. NEJM AI. 2024;1(12):AIoa2300221.

36. Liu Y, Yu W, Dillon T. Regulatory responses and approval status of artificial intelligence medical devices with a focus on China. NPJ Digit Med. 2024;7(1):255.

37. Li Z, Wang L, Wu X, Jiang J, Qiang W, Xie H, et al. Artificial intelligence in ophthalmology:

the path to the real-world clinic. Cell Rep Med. 2023;4(7):101095.

38. Ong JCL, Chang SY, William W, Butte AJ, Shah NH, Chew LST, et al. Ethical and regulatory challenges of large language models in medicine. Lancet Digit Health. 2024;6(6):e428–32.

39. Dow ER, Keenan TDL, Lad EM, Lee AY, Lee CS, Loewenstein A, et al. From data to deployment: the collaborative community on ophthalmic imaging roadmap for artificial intelligence in age-related macular degeneration. Ophthalmology. 2022;129(5):e43–59.