



## OPEN ACCESS

EDITED BY  
Santosh Kumar Sharma,  
University of Limerick, Ireland

REVIEWED BY  
Gilbert Yong San Lim,  
SingHealth, Singapore  
Luigi Di Biasi,  
University of Salerno, Italy

\*CORRESPONDENCE  
Junguo Duan  
✉ duanjg@cdutcm.edu.cn

†These authors have contributed equally to this work and share first authorship

RECEIVED 13 October 2025  
REVISED 12 December 2025  
ACCEPTED 23 December 2025  
PUBLISHED 16 January 2026

CITATION  
Liu C, Duan Y, Wu H and Duan J (2026) A panoramic perspective: application prospects and outlook of multimodal artificial intelligence in the management of diabetic retinopathy. *Front. Public Health* 13:1724001. doi: 10.3389/fpubh.2025.1724001

COPYRIGHT  
© 2026 Liu, Duan, Wu and Duan. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# A panoramic perspective: application prospects and outlook of multimodal artificial intelligence in the management of diabetic retinopathy

Chun Liu<sup>1†</sup>, Yu Duan<sup>2†</sup>, Hao Wu<sup>1</sup> and Junguo Duan<sup>1,2,3\*</sup>

<sup>1</sup>Eye College of Chengdu University of TCM, Chengdu, Sichuan, China, <sup>2</sup>Ineye Hospital of Chengdu University of TCM, Chengdu, Sichuan, China, <sup>3</sup>Eye Health with Traditional Chinese Medicine Key Laboratory of Sichuan Province, Chengdu, Sichuan, China

Diabetic retinopathy (DR) is a leading cause of blindness among the working-age population, and its management is challenged by the disease's inherent heterogeneity. Current management paradigms, based on standardized grading, are inadequate for addressing the significant inter-patient variability in disease progression and treatment response, thereby limiting the implementation of personalized medicine. While artificial intelligence (AI) has achieved breakthroughs in unimodal analysis of retinal images, the single dimension of information fails to capture the complete, complex pathophysiology of DR. Against this backdrop, multimodal AI, capable of integrating heterogeneous data from multiple sources, has garnered widespread attention and is regarded as a revolutionary tool to overcome current bottlenecks and achieve a panoramic understanding for the management of each patient. This review aims to systematically explore the frontier research and developmental potential of multimodal AI in DR management. It focuses on its data sources, core fusion technologies, and application framework across the entire management workflow. Furthermore, this review analyzes future challenges and directions, with the goal of providing a theoretical reference and guidance for the advancement of precision medicine in DR.

## KEYWORDS

artificial intelligence, data fusion, diabetic retinopathy, multimoda, precision medicine

## 1 Introduction

Diabetes mellitus (DM) has become a severe global public health challenge. According to predictions from the International Diabetes Federation, the number of patients with DM worldwide will increase to 783.2 million by 2045 (1). A commentary in *The Lancet* noted that the effectiveness of DM prevention and control over the next 20 years will profoundly impact population health and life expectancy for decades to come (2). As one of the most common microvascular complications of DM with a significant risk of blindness, diabetic retinopathy (DR) and its management are critical components of the overall DM care system. A recent meta-analysis revealed a global DR prevalence of 22.27%, with the prevalence of vision-threatening DR (VTDR) reaching 6.17% (3).

However, the impact of DR extends far beyond vision impairment alone. As a crucial marker of systemic microvascular disease, its presence is closely associated with an increased risk of severe adverse events, including cardiovascular disease, cognitive dysfunction, and premature mortality (4, 5). These comorbidities collectively pose a serious threat to an individual's quality of life, ability to work, and overall wellbeing. Consequently, implementing early diagnosis and timely treatment is not only key to halting the progression of retinopathy and preventing vision loss but also fundamental to reducing the risk of related systemic complications and improving long-term patient health outcomes (6).

Current DR management models, however, face immense challenges due to the disease's significant individual heterogeneity. Clinically, even among patients at the same stage, disease progression trajectories vary dramatically: some remain stable for long periods, while others deteriorate rapidly. Although the duration of DM and metabolic control are recognized as key risk factors, they fall short of fully explaining the vast differences in disease evolution among individuals or why some patients respond poorly to existing treatments. This inherent heterogeneity renders the "one-size-fits-all" management strategy based on traditional grading standards inadequate. Therefore, a paradigm shift from "standardized grading" to "personalized management" is urgently needed in clinical practice. The core of this shift lies in achieving precise risk stratification, dynamic disease progression prediction, and individualized treatment strategy guidance to address the complexity and uncertainty of DR (7).

Against this background, the rise of artificial intelligence (AI), particularly deep learning, presents a revolutionary opportunity to achieve efficient and precise DR management (8, 9). Among these technologies, multimodal AI has attracted widespread attention for its unique integrative analytical capabilities. It aims to construct a more comprehensive, three-dimensional, and dynamic digital model for each patient by fusing data of different types from various sources (10). This approach has the potential to profoundly characterize an individual's unique pathophysiological state and capture complex associations hidden within unimodal information, thereby providing a solid data foundation for accurate risk prediction and decision support. This represents not only a deepening of existing diagnostic capabilities but also a profound transformation toward achieving full-cycle, all-encompassing personalized management.

In view of this, this review aims to survey the frontier explorations and future potential of multimodal AI in DR management. This paper will first introduce the multimodal data sources used to construct a "panoramic perspective" of DR. Second, it will explore the core AI fusion technologies for achieving efficient data integration. Based on this, it will focus on the application potential and implementation pathways of multimodal AI across the entire DR management workflow—from early warning, precise staging, and dynamic prediction to personalized treatment decisions. Finally, this paper will analyze the challenges currently facing the field and look forward to its future directions, with the aim of providing a reference for research into intelligent DR diagnosis and treatment from a multidisciplinary perspective.

## 2 Multimodal data sources for a panoramic management of DR

The core prerequisite for achieving a "panoramic" management of DR is to move beyond the limitations of single information sources toward a comprehensive analytical framework that integrates multi-dimensional, multi-source data. The necessity of this shift is rooted in an ever-deepening understanding of DR pathophysiology: DR is no longer considered a purely vascular disease but a complex pathological process involving dysfunction of the entire neurovascular unit (11). Furthermore, a series of subclinical structural and functional changes occur in the retina long before the appearance of clinically visible microaneurysms (12). These findings collectively reveal that a truly effective management model must be able to capture and integrate these multi-level pathological changes. Therefore, constructing a comprehensive database that covers retinal macro-morphology, micro-structure, microcirculatory function, systemic risks, and even genetic background is the first step toward personalized management and the foundation upon which multimodal AI can exert its powerful analytical and predictive capabilities. This paper will systematically review the data foundations that constitute this panoramic model, categorized into ophthalmic imaging, systemic indicators, and emerging data sources.

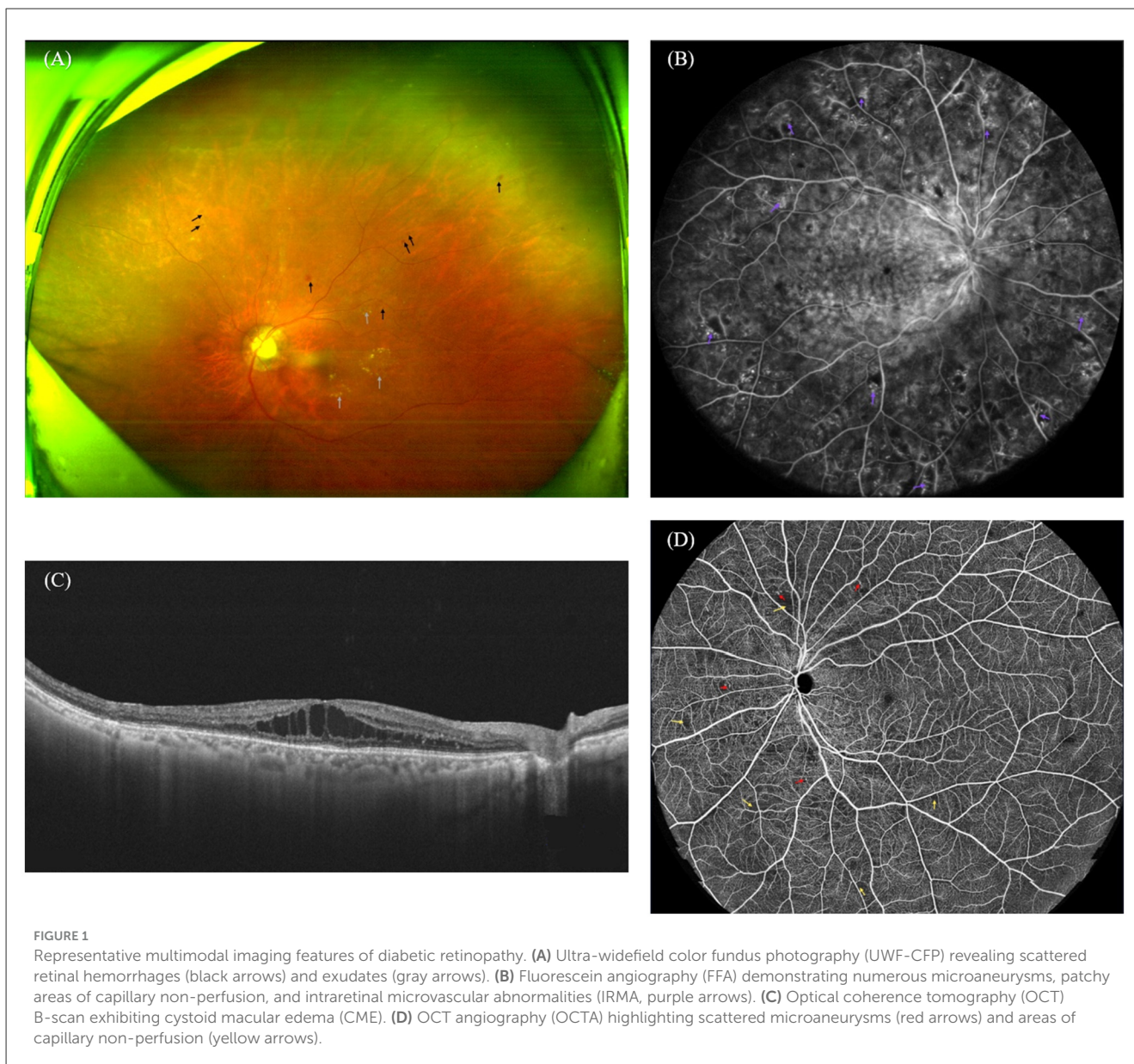
### 2.1 Ophthalmic imaging data

Ophthalmic imaging is the cornerstone of DR diagnosis, grading, and follow-up. The advancement of modern imaging technologies allows for non-invasive, high-resolution observation of the retina from multiple dimensions. The application of AI has further unlocked the potential of this imaging data, enabling the extraction of quantitative biomarkers far beyond the range of human visual recognition and facilitating an analytical leap from macro-structure to micro-function and from static assessment to dynamic prediction. A comprehensive visualization of these key pathological features across different imaging modalities is presented in [Figure 1](#).

#### 2.1.1 Color fundus photography

Color fundus photography (CFP) is the most fundamental and widely used imaging technique in DR screening and clinical research (13). By capturing a two-dimensional color image of the posterior pole of the retina, it provides a direct visualization of the classic pathological features of DR, such as microaneurysms, hemorrhages/blots, hard exudates, cotton wool spots, venous beading, intraretinal microvascular abnormalities, and neovascularization. The precise segmentation and feature extraction of these classic lesions form the basis of AI applications in DR (14).

However, traditional seven-field or single-field CFP has a significant limitation: its limited field of view, which primarily covers a 30°-50° area of the posterior pole, neglecting the expansive peripheral retina. A growing body of evidence indicates that



predominantly peripheral lesions (PPLs) are an independent and important predictor of the risk of DR progression (15). A key study found that patients with PPLs had a 3.2-fold higher risk of DR progression and a 4.7-fold higher risk of proliferative DR (PDR) than those without peripheral lesions (16). In this context, ultrawide-field (UWF) imaging technology has emerged. UWF can capture an area of up to 200° (approximately 80% of the total retinal surface area) in a single image, clearly visualizing the previously difficult-to-observe periphery and providing a powerful tool for assessing PPLs (17). Studies have shown that UWF imaging has high consistency with traditional seven-field photography within the ETDRS 7-field area, with its core added value lying in its ability to identify and quantify predominantly peripheral lesions outside the traditional fields (18).

Beyond the identification of classic lesions, the quantitative analysis of subclinical retinal geometric features provides important biomarkers for the very early detection of DR (19, 20).

Increased retinal venular caliber has been confirmed as a key early indicator. A large-scale meta-analysis confirmed that wider venular caliber is independently associated with the future risk of type 2 DM (21) and is more pronounced in patients with DR (22). Increased vascular tortuosity reflects a state of local ischemia, and recent research further points out that the tortuosity of branch retinal arteries is particularly closely associated with the genesis and severity of DR (23). As a measure of the complexity of the vascular network, a decrease in fractal dimension reflects a simplification of the vascular pattern caused by capillary occlusion. Prospective studies have confirmed that a lower baseline fractal dimension value is an independent predictor of future incident DR (24).

### 2.1.2 Fundus fluorescein angiography

For decades, fundus fluorescein angiography (FFA) has been the gold standard for evaluating retinal vasculature in

DR (25). By intravenously injecting sodium fluorescein, FFA dynamically displays the retinal vascular network with high contrast, enabling the sensitive detection of key pathological features such as microaneurysms, neovascularization, capillary non-perfusion areas, and vascular leakage. This information is crucial for the precise grading, diagnosis, and treatment decisions in DR (26, 27). FFA can clearly distinguish between intraretinal microvascular abnormalities and neovascularization and can locate the source of leakage causing macular edema to guide precise laser therapy (28).

Despite its limitations, such as its invasive nature, the rich dynamic vascular information provided by FFA (e.g., quantifiable metrics like non-perfusion area and leakage index) serves as an extremely valuable data source for AI models. Integrating these functional imaging features into AI models has the potential to surpass lesion identification based on static CFP, thereby enhancing the application potential of AI in precision DR management (29).

### 2.1.3 Optical coherence tomography

Compared to CFP, which provides a two-dimensional planar view of the retina, optical coherence tomography (OCT) presents its fine three-dimensional cross-sectional anatomy. Using the principle of low-coherence interferometry, OCT performs non-invasive tomographic scans of the retina, clearly displaying its various layers with micron-level resolution. One of its core applications in DR management is the precise quantification of diabetic macular edema (DME). DME is the leading cause of vision loss in patients with DR and is characterized by the breakdown of the blood-retinal barrier and fluid accumulation in the macular region (30). OCT can accurately measure central macular thickness and perform localization and quantitative analysis of intraretinal or subretinal fluid in the macular area (31, 32). This is not only the gold standard for DME diagnosis but also a core objective metric for evaluating the efficacy of treatments such as anti-VEGF therapy (33, 34).

Beyond quantifying DME, OCT also has an irreplaceable value in revealing early neurodegeneration in DR. Retinal neurodegeneration is considered a key event in the early pathogenesis of DR and may even precede the appearance of microvascular lesions (35). As an extension of the central nervous system, the retina provides a unique window for observing neuronal damage that may reflect systemic neurological conditions. OCT can precisely segment and measure the thickness of the retinal nerve fiber layer and the ganglion cell-inner plexiform layer complex. Changes in the thickness of these inner retinal layers are important structural parameters for early DR detection (36, 37). Numerous studies have confirmed that progressive thinning of the retinal nerve fiber layer and the ganglion cell-inner plexiform layer occurs in the early stages of DR, and even in the pre-diabetic state, and that the degree of thinning is closely related to the severity of DR and future visual function impairment (38, 39).

Furthermore, the quantitative analysis of retinal reflectivity using OCT has emerged as a novel method for revealing microstructural and compositional changes in early DR. Studies have confirmed that in DM patients without clinical DR, the reflectivity of the outer retina (especially the ellipsoid zone) is already reduced (40), while changes in the reflectivity of the

inner retina (ganglion cell layer) are also closely associated with neurodegeneration (41). These emerging OCT biomarkers provide AI models with richer, quantifiable structural information, opening up new avenues for the very early detection of DR.

### 2.1.4 OCT angiography

Optical Coherence Tomography Angiography (OCTA) represents a major technological advance in ophthalmic imaging in recent years. Building on OCT, it achieves non-invasive, layer-specific, three-dimensional visualization of retinal and choroidal blood flow by analyzing the decorrelation signal generated by the movement of blood cells between consecutive B-scans (42). Compared to traditional FFA, OCTA does not require the injection of a contrast agent, is rapid, and provides layer-specific blood flow information that is unattainable with FFA (43).

OCTA provides crucial microcirculatory functional information for a panoramic assessment of DR. In the pathophysiology of DR, capillary non-perfusion is considered a core event driving disease progression (11). OCTA can precisely delineate and quantify the size, morphology, and spatial distribution of non-perfusion areas, and its distribution characteristics show significant differences among patients with varying degrees of DR severity (44). Notably, morphological changes in the foveal avascular zone in the macula have been confirmed as powerful predictors of DR progression and visual prognosis (45, 46). Additionally, OCTA can calculate metrics such as vascular density and perfusion density in different regions and at different depths (47). Prospective studies have found that a lower baseline peripapillary vascular density is an independent predictor for the incidence of DR and progression to referable DR over 2 years (48).

## 2.2 Systemic data

As a local microvascular complication of DM in the eye, the pathophysiological process of DR is closely related to the patient's overall systemic condition. The onset and progression of DR are not only influenced by local factors but are also intricately intertwined with long-term metabolic disorders, comorbidities, and treatment adherence. Therefore, any model that detaches from the systemic context and analyzes ophthalmic images in isolation has inherent limitations, as it cannot fully characterize the complete disease state, thereby limiting the accuracy of risk prediction. To construct a “panoramic perspective” that can accurately reflect the full picture of the disease, incorporating systemic data into the analytical framework is a necessary prerequisite for achieving a comprehensive and dynamic assessment.

### 2.2.1 Electronic health records

The rich, longitudinal clinical information contained in Electronic Health Records (EHRs) serves as a critical bridge connecting ocular pathology with systemic status and constructing a “panoramic perspective.” Among the numerous systemic data types, several categories are particularly crucial for the precise management of DR. First, glycemic control indicators, represented

by glycated hemoglobin as the gold standard for assessing long-term blood sugar levels, have been confirmed as the strongest clinical risk factor for the onset and progression of DR (49). Second, metabolic indicators such as blood pressure and lipids are also recognized as independent risk factors; effective control of hypertension and hyperlipidemia can significantly delay the progression of DR (50). Furthermore, the duration and type of DM provide a baseline risk context, with a longer duration typically correlating with a higher prevalence and severity of DR (51), and with type 1 and type 2 DM exhibiting differences in DR presentation and complications (52). Concurrently, systemic comorbidities, especially diabetic kidney disease, which shares a highly homologous microvascular pathogenesis with DR, often herald an accelerated deterioration of DR (53). Finally, the patient's medication history, including the use of glucose-lowering, antihypertensive, and lipid-lowering drugs, as well as insulin, directly reflects their treatment status and adherence, making it an indispensable component for building dynamic risk models (54).

This structured and unstructured data embedded within EHRs can be efficiently extracted and integrated using AI techniques such as natural language processing, providing high-quality input for multimodal fusion and thereby significantly enhancing the predictive performance of the models. In terms of data storage and retrieval infrastructure, while laboratory data is initially generated in specialized Laboratory Information Systems (LIS), it is typically integrated into the central EHR system via interoperability standards to allow for unified access. Consequently, the EHR and LIS function as distinct but interconnected databases, with the EHR serving as the primary interface for clinical decision-making. However, this integration process may lead to data redundancy, such as duplicates arising from both automated interface transmission and manual clinical entry. To address potential inconsistencies between these sources, data preprocessing pipelines for DR management models must employ rigorous cleaning strategies. These typically involve designating the direct LIS feed as the "source of truth" or applying timestamp-based logic to prioritize the most recent and accurate values, ensuring the fidelity of the data fed into the AI models.

## 2.2.2 Laboratory data

In addition to routine clinical indicators, laboratory data can provide supplementary information for DR risk assessment. These metrics reflect the impact of the systemic metabolic state on retinal pathology and represent potential novel biomarkers (55). Among them, renal function indicators are particularly closely associated with DR. Multiple studies have confirmed that blood urea nitrogen is a significant risk factor for DR (56), with elevated levels, especially above 20 mg/dl, significantly increasing the risk of DR (57). Serum creatinine has also been identified as a key indicator in several predictive models (58). Moreover, urinary microalbumin and the serum uric acid to creatinine ratio have shown significant correlations with DR. Notably, some research has found that in certain patient populations, a higher estimated glomerular filtration rate is also associated with an increased risk of DR (59).

In recent years, research on inflammation-related hematological indicators has offered a new perspective on

DR risk prediction, with composite markers such as the systemic immune-inflammation index (SII) and systemic inflammatory response index (SIRI) being particularly prominent. Studies have confirmed that SII is a sensitive indicator for predicting complications and mortality in patients with type 2 DR (60). Furthermore, both SII and SIRI are independent risk factors for DR, and their combined use can significantly improve diagnostic accuracy (AUC = 0.782), showing great potential in the diagnosis and stratified management of DR (61). Crucially, studies have shown that levels of NLR, SII, and SIRI are already significantly elevated before the appearance of clinically visible DR lesions and are negatively correlated with subclinical microvascular damage in the retina detected by SS-OCTA (62). As DR progresses, these inflammatory markers also show a stepwise increase, with SII demonstrating a significant linear relationship with the onset and development of DR, serving as an independent risk factor at all stages of the disease (63).

In summary, laboratory data not only provide important biological explanations for the onset and progression of DR but also serve as a key input for AI models. By quantifying systemic renal function and inflammatory status, they help to achieve early identification and individualized prediction of DR risk.

## 2.3 Emerging data sources

With the advent of the precision medicine era, "omics" data—represented by genomics, proteomics, and metabolomics—are gradually becoming a new dimension for constructing a panoramic view of disease. As previously mentioned, significant individual heterogeneity exists in the onset and progression of DR, with genetic susceptibility considered an important contributing factor, accounting for as much as 25%–50% of the risk of DR onset (64). Genome-wide association studies have identified multiple DR susceptibility loci, but their associations often exhibit ethnic specificity. For example, a Brazilian study found that specific polymorphisms in the TIE2 and ANGPT-1 genes are associated with protection against DR (65), whereas new susceptibility genes identified in a Japanese study, STT3B and PALM2, have not been validated in other populations (64). Similarly, a variation in the PRMT1 gene was found to be associated with an increased incidence of PDR in Japanese patients with type 2 DM (66). Meta-analyses have further consolidated these findings, clarifying that polymorphisms in cytokine genes (such as IFN- $\gamma$  rs2430561 and TGF- $\beta$  rs1800471) are associated with a reduced susceptibility to DR (67), and that specific SNPs (such as rs2010963, rs833061) are associated with the onset and progression of NPDR and PDR (68). A recent review pointed out that the genetic polymorphisms of DR are mainly concentrated in key genes such as VEGF, ACE, and APOE (69). Integrating this information into a genetic risk score can serve as an independent data modality for AI models, helping to explain differences in individual prognoses under similar clinical conditions.

In contrast to the static perspective of genomics in revealing genetic susceptibility, proteomics and metabolomics can provide a dynamic snapshot of the disease state at the molecular level, making them of great interest as they are closer to the actual

disease phenotype. In the field of metabolomics, researchers are dedicated to finding characteristic metabolites that reflect the onset and progression of DR. Progress has been made in studies on easily accessible biological samples such as serum: a cross-sectional study found that reduced levels of circulating L-Tyrosine can serve as an indicator of the presence of retinopathy in patients with T2DM (70); while a large-scale longitudinal study identified 17 metabolites associated with the risk of incident DR, noting that various N-lactoyl amino acids increase the risk, whereas citrulline is associated with a reduced risk (71). In addition to blood samples, research has also extended to non-invasive specimens like tear fluid, with one study finding that elevated levels of lactate in the tears of PDR patients can serve as an independent risk factor for assessing PDR (72). At the same time, comprehensive studies have revealed dysregulation in multiple pathways in patients with DR, including arginine-proline metabolism, amino acid metabolism, and purine metabolism. These findings provide important clues for elucidating the pathological mechanisms of DR and exploring new therapeutic targets (73).

In the field of proteomics, analysis of intraocular fluids can provide direct insights into the local pathophysiology of the retina. As the vitreous humor is directly adjacent to the retina, changes in its proteome can effectively reflect changes in retinal homeostasis. Studies have shown that proteins related to coagulation, the complement system, and the kallikrein-kinin system are significantly upregulated in the vitreous humor of patients with PDR and can serve as potential biomarkers (74). Proteomic analysis of the aqueous humor has also identified a large number of differentially expressed proteins associated with processes such as inflammation, oxidative stress, and apoptosis, including apolipoprotein A-I, selenoprotein P, and cystathionine  $\beta$ -synthase (75). These findings collectively confirm the complex pathophysiological network of DR, involving local immune inflammation, angiogenesis, coagulation dysfunction, and tissue repair (76).

Although the acquisition costs and analytical complexity of “omics” data are currently high, they represent the future direction of personalized medicine. The multimodal fusion of this deep molecular information with clinical and imaging data has the potential to enable AI models to break through existing bottlenecks and achieve a more precise prediction of DR risk.

### 3 Key technologies and methodologies of multimodal AI

After constructing a multimodal database encompassing ophthalmic imaging, systemic clinical information, unstructured medical texts, and emerging “omics” data, the core challenge lies in effectively integrating these data, which vary in origin, complexity, and dimensionality. The goal is to achieve informational complementarity and synergy, ultimately enhancing the characterization of the disease state and the assessment of clinical risk. Multimodal fusion is a key approach to solving this problem. In recent years, the development of relevant computational methods has provided new ideas and tools for this

field, gradually making DR management based on multimodal data a possibility.

## 3.1 Overview of fusion strategies

The fundamental purpose of multimodal fusion is to establish a computational framework that can collaboratively analyze information from different sources to improve the accuracy and reliability of specific clinical tasks. Based on the stage at which information is integrated within the model, classic fusion strategies can be divided into early fusion, late fusion, and hybrid fusion (77). Beyond these traditional architectures, the integration of Large Language Models (LLMs) represents a cutting-edge evolution in multimodal analysis, particularly for processing unstructured text-image pairs, introducing a new paradigm for clinical decision support.

### 3.1.1 Early fusion

Early fusion, also known as feature-level fusion, is the most direct strategy for multimodal integration. The basic idea is to concatenate or combine features from different modalities at the initial stage of analysis to form a unified high-dimensional feature vector, which is then fed into a downstream predictive model for training (78). For instance, in the CAD-EYE system developed by Khalid et al., the authors employed an early fusion approach by combining feature vectors extracted from two distinct deep learning architectures, MobileNet and EfficientNet. This integration of features at the encoding stage allowed the model to achieve a high classification accuracy of 98% for multi-eye disease diagnosis, surpassing single-model baselines (79). Similarly, Ejaz et al. utilized a parallel CNN framework where deep features extracted from fundus images were fused using Canonical Correlation Analysis before the classification step. This feature-level integration enabled the model to capture more discriminative patterns, resulting in a detection accuracy of 93.39% (80). Hervella et al., through a self-supervised pre-training method, enabled a model to simultaneously learn both common and unique features between multimodal images, thereby constructing a powerful unified feature encoder. This encoder performed exceptionally well in the subsequent DR grading task, validating the great potential of the early fusion strategy in enhancing the model's generalization ability for downstream tasks and addressing the issue of sparse data annotation (81). The advantage of this method is its ability to integrate information from different modalities at the outset of modeling, thereby capturing potential cross-modal correlations and interaction patterns, which helps to improve model performance.

However, its limitations should not be overlooked. First, differences in time scales, spatial resolutions, and data structures across modalities make data alignment difficult, and simple concatenation may introduce noise or cause information loss. Second, discrepancies in feature scales can lead to high-dimensional modalities (such as imaging data) dominating the model, thereby overshadowing important information from other

modalities. Finally, this strategy is highly dependent on data completeness; if one modality is missing, the stability and applicability of the overall model will be affected.

### 3.1.2 Late fusion

Late fusion, also known as decision-level fusion, operates opposite to early fusion. The approach involves first building separate, independent models for each modality, each of which outputs a prediction or decision score based on its own data. Subsequently, at the final stage of analysis, these outputs are integrated using methods such as weighted averaging, majority voting, or a small meta-learner to arrive at a final conclusion (82). A classic application of this strategy is the ensemble approach proposed by Qummar et al., where five independent CNNs—ResNet50, InceptionV3, Xception, Dense121, and Dense169—were trained on fundus images. The final DR detection result was derived by calculating the weighted average of the probability scores from these distinct models, yielding higher robustness than any single model (83). Furthermore, a representative application of this strategy is the DRCNN-Lesion Proxy framework proposed by Sekar et al. In this architecture, global image-level features extracted by a ResNet34 backbone and lesion-specific cues simulated by a proxy module are processed independently and then integrated through a late fusion classification head. By fusing these heterogeneous information sources at the final decision stage, the model achieved robust DR severity prediction with an accuracy of up to 98.37% across multiple public datasets (84). The advantage of this strategy lies in its greater flexibility and modularity. Researchers can select the most appropriate analytical method for each modality based on its characteristics, and the system can still operate on the remaining modalities when some data are incomplete, thus exhibiting good robustness. However, its limitation is that information is integrated only at the decision level, ignoring potential interactions and synergies between modalities during the feature-learning stage. For example, the pathological link between retinal microvascular morphology and neural layer thickness often cannot be effectively utilized in a late fusion framework, thereby limiting the model's ability to uncover deeper patterns of the disease mechanism.

### 3.1.3 Hybrid/deep fusion

To combine the deep interaction of early fusion with the flexibility of late fusion, hybrid or deep fusion strategies have become a frontier in multimodal research. The core idea of these methods is to abandon a “one-off” integration at the model's input or output, and instead achieve multiple, in-depth information exchanges at the intermediate layers of the model (82). A typical architecture usually involves building separate encoders for each modality or data source to learn their respective feature representations, with connections established between different layers of the networks to enable the dynamic flow and deep integration of cross-modal information. This paradigm aims to introduce cross-modal collaboration at the early stages of feature learning while preserving the independence and specificity of each modality's processing pathway, with the goal of achieving optimal performance.

The superiority of hybrid fusion has been demonstrated in DR research. When processing OCTA data from different retinal depths, a study by Ebrahimi et al. (85) clearly indicated that, compared to early fusion, late fusion, and single-layer inputs, an “intermediate fusion” architecture that fuses features from the superficial, deep, and choriocapillaris layers at the network's middle layers could increase DR classification accuracy to 92.65%, achieving the best performance. Similarly, to integrate the complementary information from different fields of view in OCTA, Li et al. designed a hybrid fusion framework to jointly analyze high-resolution ( $6 \times 6 \text{ mm}^2$ ) and ultrawide-field ( $15 \times 15 \text{ mm}^2$ ) OCTA images. The results showed that the performance of this fusion strategy in detecting all stages of DR (including early and late) was significantly superior to algorithms using only a single field of view, proving the effectiveness of fusing local fine structures with global vascular layout information (86). Furthermore, the design of hybrid architectures is becoming increasingly sophisticated. Tseng et al. proposed a “two-stage early fusion” model that mimics the diagnostic workflow of an ophthalmologist. The model first performs lesion detection and then severity grading. This sequential hybrid method identified early DR more accurately than traditional algorithms, demonstrating the potential of hybrid fusion in enhancing model robustness and credibility (87).

In practice, achieving the aforementioned deep fusion relies on several advanced computational methods, with the attention mechanism, Transformer, and Graph Neural Network (GNN) being the most representative. The attention mechanism, through cross-modal query, allows a model to dynamically and selectively focus on relevant regions in one modality (e.g., fundus images) based on features from another modality (e.g., high-risk indicators in EHRs), achieving intelligent information weighting (88). The Transformer, with its powerful self-attention mechanism, can effectively capture and fuse long-range dependencies in both time-series and image-series data, thereby accurately characterizing the dynamic evolution of the disease (89). Some studies have attempted to fuse the local feature extraction capabilities of Convolutional Neural Networks (CNNs) with the global modeling capabilities of Vision Transformers (ViTs), achieving high-precision prediction for DR grading (90). Meanwhile, the GNN offers a structured fusion perspective. It models data from different modalities as nodes in a graph and learns their intrinsic connections through a message-passing mechanism. This allows for the systematic characterization of a “panoramic patient profile” that includes imaging, clinical, and “omics” information, thereby generating higher-level disease representations.

### 3.1.4 Large language models and text-image integration

The integration of LLMs and Vision-Language Models represents a significant leap in processing the unstructured textual component of multimodal data. Unlike traditional NLP methods that rely on rigid rule-based extraction, LLMs demonstrate a superior ability to understand complex clinical narratives and facilitate conversational decision support. However, their direct application in high-stakes medical diagnosis requires critical validation against established deep learning baselines.

Current evidence suggests that while LLMs excel in interaction, they may struggle with precision in specialized diagnostic tasks compared to dedicated models. For instance, a recent study by Rossi et al. compared a WE-LSTM (Word Embedding-Long Short-Term Memory) network against a WizardLM-powered chatbot (DiabeTalk) for diabetes diagnosis. The results revealed a stark contrast: while the LLM-based chatbot offered a user-friendly conversational interface and “acceptable results” without specific training, the specialized WE-LSTM model significantly outperformed it in diagnostic accuracy (97.80 vs. 77.56%) and specificity (95.90 vs. 57.80%) when applied to minimally pre-processed data (91).

This highlights a crucial limitation in current multimodal frameworks: general-purpose foundation models exhibit strong logic and language understanding, but they cannot yet replace specialized, fine-tuned predictive models for precision tasks without extensive domain-specific adaptation. Therefore, the current frontier involves integrating the reasoning capabilities of LLMs with the precision of specialized architectures (like CNNs for images or LSTMs for sequential clinical data) to create systems that are both accurate and communicatively competent.

## 3.2 Explainable AI (XAI)

Although multimodal AI models show great potential in the management of DR, their clinical translation still faces a key challenge: the inherent “black box” nature of deep learning models (92). In a high-risk field like medicine, an AI system that cannot clearly articulate its internal decision-making logic is unlikely to gain the full trust and adoption of clinicians. Therefore, the transparency and interpretability of the decision-making process are necessary prerequisites for promoting the widespread deployment of AI models in clinical settings and realizing their applied value (93). The development and application of XAI technologies, which aim to open the “black box” and clarify the contribution weights of key modalities and features, are a crucial step in transforming multimodal AI from a “high-performance prediction tool” to a “trustworthy clinical decision partner.”

### 3.2.1 Explanation for visual modalities

When processing core visual modalities like fundus images, explanation methods based on Class Activation Mapping are the most widely used, with Gradient-weighted Class Activation Mapping and its variants becoming mainstream tools. These techniques generate “saliency heatmaps” that intuitively highlight the key image regions upon which a convolutional neural network relies for DR grading or lesion segmentation. Sharma et al. (94) used Grad-CAM to verify that their model was indeed focusing on true pathological areas; Rautaray et al. (95) employed Grad-CAM++ to clearly display the key retinal structures that influenced DR severity grading. Such visual validation methods provide clinicians with an effective review tool, allowing them to quickly determine whether a model’s judgment is based on true pathological features like microaneurysms and exudates, thereby enhancing trust in the model’s diagnostic logic (96).

More importantly, the value of XAI extends far beyond validation to aiding scientific discovery. A groundbreaking study used Guided Grad-CAM technology to explore sex differences in DR, finding that the model focused more on the macular region when identifying female patients and more on the optic disc and peripheral vessels for male patients. This discovery led to a new clinical hypothesis: female DR patients may be more prone to developing macular edema, while males face a higher risk of proliferative DR (PDR) (97). This fully demonstrates that XAI can not only “explain AI” but also inspire new human understanding of the disease itself.

### 3.2.2 Explanation for multimodal fusion

For complex models that fuse multimodal data such as fundus images, clinical indicators, and “omics” information, model-agnostic XAI methods offer more powerful explanatory capabilities, capable of revealing the interactions between different modalities. Local Interpretable Model-agnostic Explanations (LIME) approximates the local behavior of a complex model with a simple one by making perturbations around a single sample, thereby revealing the prediction mechanism for that sample (98). In a multimodal context, LIME can clearly quantify the contribution of different data modalities and their internal features to the prediction for an individual patient, providing an intuitive basis for personalized decision-making.

Meanwhile, the game theory-based SHAP (SHapley Additive exPlanations) method has been widely used in recent multimodal DR research due to its ability to provide both local and global explanations and its solid theoretical foundation. For each individualized prediction, SHAP can precisely calculate the marginal contribution of each input feature to the final prediction, clearly revealing the key factors driving individual risk (99). Recent studies have fully demonstrated the great potential of SHAP in DR biomarker discovery: Zong et al. (100) used SHAP to explain their XGBoost model and successfully identified various metabolites (such as C18:1OH, threonine) associated with DR risk and their risk cutoff values. Another study, also applying SHAP, found that glucose, glycine, and age were important predictors across all stages of DR, while creatinine and various phosphatidylcholines showed higher importance in the late PDR stage, suggesting they could be potential biomarkers for severe DR (101). Gui et al. combined machine learning with SHAP to explore the relationship between heavy metal exposure and DR. Their analysis clearly indicated that, among numerous variables, urinary antimony level was the most critical factor influencing predicted DR risk, with a contribution weight far exceeding other variables. This finding not only reveals the potential impact of environmental pollutants on DR but also provides a new direction for early non-invasive screening (102). These cases strongly demonstrate that methods like SHAP can help researchers accurately trace the source from high-dimensional, complex multimodal data to locate key driving factors, with their value extending from mere “model explanation” to “aiding knowledge discovery.”

In general, XAI is a core technology that empowers the clinical translation of multimodal AI in DR management. Its value is reflected on three levels: first, building clinical trust by ensuring that

the AI's decision-making process can be reviewed and understood through transparent, intuitive explanations; second, deepening disease understanding by helping researchers gain insights into potential pathophysiological mechanisms from complex data correlations; and third, driving scientific discovery by serving as an efficient biomarker discovery engine to identify new risk factors from high-throughput data.

The organic combination of a high-performance multimodal fusion model and a powerful XAI explanation system will together form the core of the next-generation intelligent management platform for DR. Such a platform will not only provide accurate diagnoses and risk predictions but also clearly reveal why, providing solid and powerful support for achieving truly precise and personalized clinical decisions.

## 4 Frontier explorations and future directions of multimodal AI in DR management

As a complex microvascular complication, the management of DR is a continuous process that includes risk prediction, diagnostic staging, progression monitoring, treatment decision-making, and long-term follow-up (103). By integrating data from different sources, multimodal methods provide new possibilities for establishing a precise and individualized decision support system covering the entire course of DR management (77). As described below, multimodal AI has significant application potential in five key stages of DR management.

### 4.1 Precise risk stratification and early warning

The success of DR management is critically dependent on the early stages (104). Therefore, the primary step in DR management is to accurately identify high-risk individuals before the onset of clinical signs to implement proactive prevention and intervention. Traditional risk models often rely on a few clinical variables and struggle to capture complex individual risks. By integrating data from various sources, multimodal AI can construct more comprehensive and dynamic risk prediction models, achieving a paradigm shift from “disease detection” to “risk prediction.”

A highly representative direction is the fusion of EHR data with fundus images. EHRs contain long-term patient data on blood glucose, blood pressure, laboratory tests, and medication history, which constitute a time-series reflecting their systemic health status. One study combining fundus images with heterogeneous EHR data showed that its fusion model significantly outperformed single-source models in the task of screening for referable DR (AUC of 97.96%), demonstrating that fusing multi-source data can lead to earlier and more accurate referral decisions (105). Another study (106) delved deeper by utilizing 20 years of EHR data to create thousands of features capturing the dynamic evolution of patient health. The multimodal model they built showed outstanding performance in early DR detection (AUC of 0.988), confirming that the systemic health trajectory recorded in EHRs contains rich

predictive information even before identifiable lesions appear in the fundus.

Another key strategy is the fusion of multiple ophthalmic imaging modalities to enhance screening efficacy. Liu et al. evaluated the effect of adding self-imaging OCT to traditional FP. The results showed that the combined FP and SI-OCT strategy was significantly superior in both sensitivity (87.69%) and specificity (98.29%) for detecting DME compared to FP alone. More importantly, a cost-effectiveness analysis confirmed that this combined approach is highly advantageous economically, providing reliable support for the early detection and precise referral of DR (107).

To effectively mine the complex relationships between different risk factors, researchers have also developed more advanced multimodal fusion architectures. For example, the VisionTrack framework innovatively introduces a Graph Neural Network to process clinical risk factors, treating different factors as nodes in a graph to learn their potential high-order, non-linear interactions. The framework further integrates a Convolutional Neural Network for processing images and a large language model for parsing clinical notes, significantly improving prediction accuracy by fusing diverse data sources and providing a more comprehensive assessment of retinal health (108).

### 4.2 Comprehensive diagnosis and refined staging

The accurate diagnosis and refined staging of DR are the cornerstones for formulating subsequent treatment and management strategies (109). Multimodal data fusion is a core strategy for enhancing the diagnostic performance of AI, and its advantages have been confirmed in studies integrating various types of imaging, clinical, and even biological sample information. At the imaging level, studies have confirmed that combining multiple imaging modalities can provide complementary pathological perspectives (110). For example, one study that simultaneously analyzed CFP and FFA and extracted key quantitative features using a curvelet transform built an SVM classifier that achieved 100% sensitivity and specificity in a three-level DR staging task (111). At the level of fusing imaging and clinical data, Sandhu et al. evaluated the diagnostic performance for NPDR by fusing clinical data, OCT, and OCTA. The results showed that after integrating clinical data, the model's accuracy reached 96% and its AUC reached 0.96, far surpassing the performance of any single modality (112). Even more innovatively, research has combined proteomics data from non-invasive biological samples (such as tear fluid) with fundus images, demonstrating that this cross-disciplinary data fusion can further improve the sensitivity and specificity of DR diagnosis (113).

To effectively utilize multimodal data, researchers have developed various advanced AI methods. To address the challenge of a scarcity of large-scale annotated data in clinical settings, self-supervised learning has become an important breakthrough. Some research has used multimodal data like fundus images and FFA for self-supervised feature learning, achieving diagnostic performance comparable to supervised models and providing an

effective solution for data-scarce scenarios (114). Hervella et al. (81) developed a self-supervised pre-training method that significantly improved the accuracy of the subsequent DR severity grading task by learning both the common and unique features between different modalities.

In terms of model architecture, researchers have designed various sophisticated networks to achieve efficient fusion. For example, a modality-specific attention network designed specific attention modules for CFP and OCT images to learn their complementary information (115). The TFA-Net model, through a twofold feature augmentation mechanism, effectively fused CFP with wide-field SS-OCTA images, showing excellent performance on small datasets (116). In addition, a multimodal information bottleneck network utilizes multicolor imaging technology to simultaneously extract features from multiple modalities to improve DR detection accuracy (117).

As a key component of refined staging, retinal vessel segmentation also benefits from multimodal AI, effectively overcoming the challenges faced by traditional methods (118). For example, CMFNet effectively solved the problem of discontinuous microvessel segmentation by fusing 3D volumetric data from OCTA with 2D projection maps (119). M3B-Net utilized the richer information from FFA images to assist and improve the accuracy of vessel segmentation on UWF images (120). The ELEMENT method innovatively used vessel connectivity as a key feature for machine learning classification, effectively reducing segmentation inconsistencies (121).

Finally, a systematic review and meta-analysis covering 47 studies ultimately confirmed that deep learning-based multimodal methods demonstrate high accuracy in DR detection and have the potential to be reliable automated diagnostic tools in the clinic (122). In summary, through data fusion, methodological innovation, and the application of key technologies, multimodal AI is driving the diagnosis and staging of DR toward greater precision and comprehensiveness.

### 4.3 Dynamic disease progression prediction

DR is a chronic, progressive disease. Accurately predicting the probability and time window for a specific patient to progress from non-proliferative DR (NPDR) to PDR or to develop DME is crucial for seizing the optimal moment for intervention and preventing irreversible vision loss (123). However, static, single-time-point examinations are insufficient to capture the dynamic evolution of the disease. Multimodal AI, especially models capable of processing time-series data, offers a possible path toward achieving dynamic disease progression prediction (124).

Currently, researchers have explored and validated the potential of AI in predicting DR progression from multiple dimensions. At the clinical imaging level, studies have confirmed that deep learning models can not only predict DR progression over 2 years with high accuracy using only baseline fundus images (125), but can also accurately predict disease evolution over the next 5 years by analyzing a series of retinal images in multi-ethnic datasets (126).

At the molecular biology level, research has begun to delve into the predictive information within gene expression data. For

example, one study analyzed single-cell transcriptomics data from the fibrovascular membranes of PDR patients, identified differential gene expression patterns in specific cells, and built a high-accuracy machine learning prediction model based on these findings (AUC of 0.83–0.96), providing a new avenue for predicting PDR risk from a molecular dimension (127).

Furthermore, some cutting-edge research is dedicated to visualizing the future evolution of the disease using AI. For example, the DRForecastGAN framework uses generative adversarial network technology to synthesize potential future fundus images based on a patient's current images. The model shows superior accuracy in predicting future DR severity (AUC of 0.85) and can intuitively “depict” the progression of the disease, providing an innovative visual tool for patient-doctor communication and treatment decisions (128).

To integrate data from the aforementioned different sources and dimensions to build more powerful predictive models, researchers are exploring more advanced fusion architectures. The Transformer is considered an ideal choice for this task due to its outstanding ability to process sequential data and capture long-range dependencies, making it suitable for fusing longitudinal EHR data, series of images, and even molecular biomarkers. A Transformer-based multimodal model can not only analyze cross-modal associations at each point in time but also learn the disease's evolution pattern along the entire timeline, thereby providing clinicians with a dynamic, forward-looking view for disease management.

### 4.4 Personalized treatment decision support

With the popularization of treatments such as anti-VEGF drugs, a new era has begun for the treatment of DR. However, there is significant heterogeneity in how different patients respond to the same treatment regimen. Therefore, accurately predicting a patient's response to a specific treatment is key to achieving personalized medicine (129). By deeply mining pre-treatment data, multimodal AI has the potential to identify biomarkers that are predictive of treatment response.

Imaging biomarkers are an important basis for predicting treatment response. One study using quantitative UWFA found that baseline vascular leakage patterns could predict the speed and duration of the response to anti-VEGF therapy: lower diffuse leakage predicted a faster response, while a higher ratio of perivascular to diffuse leakage predicted a more lasting effect (130). In addition, the RetmarkerDR model, by calculating the dynamic imaging metric of “microaneurysm turnover,” can not only assist in judging DR progression but also be used to evaluate treatment efficacy, providing a new tool for quantifying therapeutic effects (131).

Molecular biomarkers provide deeper biological information for prediction. A pioneering study integrated metabolomics and lipidomics data from aqueous humor samples and, combined with machine learning, successfully identified metabolites that could effectively predict a strong therapeutic response. The model they built can accurately screen for potential weak responders

(132). Similarly, by analyzing the angiogenic properties of specific proteins (such as decorin) in the aqueous humor, the efficacy of anti-VEGF drugs can also be predicted (133).

Looking to the future, an ideal personalized treatment decision model for DR will integrate baseline imaging, molecular biomarkers, and patient EHR and genomics data. By training on large data cohorts and combining with XAI techniques, the model will not only be able to predict efficacy but also explain the reasons, thereby providing clinicians with direct and actionable decision support.

## 4.5 Intelligent follow-up interval recommendation

Optimizing follow-up management is a core issue in the long-term care of DR, directly related to the effective use of medical resources and the patient's visual prognosis (109). Current follow-up intervals are mostly based on static disease staging guidelines and do not fully consider the individualized risk of the patient. This can lead to the over-treatment of low-risk, stable patients and insufficient follow-up for high-risk, progressive patients (134). By conducting a comprehensive and dynamic risk assessment of the patient, multimodal AI can provide data-driven support for recommending personalized follow-up intervals, achieving an optimal allocation of medical resources.

An intelligent follow-up recommendation system is essentially a top-level application built on the aforementioned risk stratification and progression prediction models. The DeepDR Plus system is a prime example; its research showed that through AI-based personalized risk assessment, the average screening interval could be extended from the standard 12–31.97 months, while reducing the delayed detection rate of DR progression to 0.18% (135). This strongly demonstrates the feasibility and superiority of dynamically adjusting screening intervals. Based on such evidence, some studies have even suggested that for T2DM patients without DR, the screening interval could be safely extended to 2–3 years (134).

The decision logic of such intelligent systems is dynamic and multidimensional: first, the system will conduct an initial risk assessment based on a risk stratification model and recommend a longer initial follow-up interval for low-risk individuals. Second, at each follow-up visit, the system will use a dynamic progression prediction model, inputting the latest multimodal data to update the patient's probability of future worsening events. If the probability significantly increases, the system will automatically issue an alert and recommend shortening the follow-up interval. Finally, for patients undergoing treatment, the aforementioned treatment response model can evaluate the efficacy of the treatment and adjust the follow-up plan accordingly.

In this way, multimodal AI transforms DR follow-up management from a population-based, static model to an individual-based, dynamic, and adaptive one. This model not only ensures that high-risk patients receive the most timely monitoring and intervention, preventing vision loss due to delays, but also significantly improves the efficiency and cost-effectiveness of the

healthcare system by reducing the unnecessary clinical burden on low-risk, stable patients, ultimately achieving a win-win for both patient benefit and medical resource optimization (136).

## 4.6 Critical assessment: clinical feasibility vs. experimental frontiers

While the potential of multimodal AI is vast, a critical distinction must be drawn between technologies that are currently feasible for clinical deployment and those that remain in the experimental “proof-of-concept” phase. Currently, unimodal AI systems based on fundus photography have reached a high level of maturity, with several achieving regulatory approval and real-world implementation. These systems benefit from standardized imaging protocols, clear ground truths, and established reimbursement pathways.

In contrast, the “panoramic” multimodal frameworks discussed in this review—particularly those integrating multi-omics, longitudinal EHRs, and foundation models (LLMs)—are primarily experimental. As evidenced by the comparison between WE-LSTM and WizardLM, the integration of LLMs for clinical decision support faces significant hurdles regarding “hallucinations,” lack of interpretability, and a performance gap compared to specialized networks in structured tasks. Furthermore, the infrastructure required to unify disparate data sources in real-time does not yet exist in most healthcare settings (91). Therefore, while multimodal fusion represents the theoretical future of precision medicine, its immediate clinical utility is currently limited to research hospitals with integrated data lakes, whereas widespread adoption awaits the resolution of interoperability standards and rigorous prospective validation of these complex, heterogeneous models.

## 5 Challenges faced

Although multimodal AI shows great potential in the “panoramic” management of DR, successfully translating it from theoretical research into a widely used clinical tool still requires overcoming multiple challenges at the data, technical, and application levels (103, 137). Clearly understanding these challenges and planning future directions accordingly is crucial for promoting the healthy and sustainable development of this field.

### 5.1 Data-level challenges

High-quality, large-scale, and standardized multimodal data are the core driving force behind the development of AI models. However, there are currently three major bottlenecks at the data level (138). First is the issue of data silos and sharing difficulties. Multimodal data from DR patients are often stored dispersedly in different hospitals, departments, and even different information systems, forming hard-to-surmount “data silos.” Strict patient privacy regulations and institutional data barriers make large-scale, multi-center data integration extremely difficult. The lack

of effective data sharing mechanisms greatly limits the scale and diversity of data needed for model training, which is one of the most common limitations in current research (139). Second is the problem of standardizing multi-center data. Different medical institutions use different imaging equipment, acquisition protocols, image quality standards, and formats and terminologies for recording clinical data (140). This heterogeneity leads to a severe data distribution shift, causing models trained at one center or in a specific clinical trial to often fail to achieve the same excellent performance on data from another center or in the real world, which severely affects the model's generalization ability (141). Finally, there is a lack of high-quality public datasets. Currently, most datasets used for multimodal ophthalmic AI research are limited in scale and are often private (142). As numerous literature reviews have pointed out, there is a general lack of large-scale public datasets freely available to researchers in the field of ophthalmic AI (143). There is a particular scarcity of large-scale, public, multimodal datasets that include long-term follow-up information in longitudinal cohorts. While cross-sectional data can be used to develop diagnostic models, longitudinal data are indispensable for the crucial task of "progression prediction" in DR management (144).

## 5.2 Technical-level challenges

At the algorithm and model level, there are also difficult hurdles to overcome. The first is the generalization ability and robustness of the models. As mentioned earlier, due to data heterogeneity, many AI models are at risk of "overfitting," performing excellently on the training dataset but experiencing a sharp decline in performance when encountering low-quality images, rare lesions, or different ethnic populations in a real clinical environment (145). Developing robust models that can maintain stable and reliable performance in complex and variable clinical scenarios is a key focus of current technical research. The second is the limitation of causal inference. Current deep learning models are essentially powerful "pattern recognizers," excelling at discovering complex correlations in data, but not causality. Prediction models based on correlation pose potential risks when guiding clinical interventions and therefore must be interpreted with caution and validated with prospective studies. Finally, there is the need for the deepening of XAI models. Although techniques like Grad-CAM provide preliminary tools for visual explanation, they are mostly focused on explaining a single modality. For multimodal models, we not only need to know which region of an image the model is focusing on but also understand how it balances and integrates information from different modalities. Developing XAI methods that can clearly reveal the cross-modal decision-making logic is key to building clinical trust and promoting the true integration of models into the decision-making process (146).

## 5.3 Application-level challenges

Successfully deploying multimodal AI into clinical practice still faces numerous real-world obstacles. First is the challenge

of seamless integration with clinical workflows. Busy clinical workflows have extremely high demands for efficiency. An AI system that requires doctors to manually upload multiple types of data, is complex to operate, and has long waiting times is unlikely to be accepted (147). Designing a user-friendly system that can automatically pull data from hospital information systems, run analyses in the background, and present results in a simple and intuitive way is key to determining whether the technology can be implemented. Second is the need for strict regulatory approval. As medical devices, AI software must undergo rigorous approval processes by agencies such as the National Medical Products Administration or the U.S. Food and Drug Administration (148). Compared to unimodal diagnostic support software, the validation and approval process for multimodal AI systems used for prediction and management is more complex and takes longer. Furthermore, there is the issue of cost-effectiveness analysis. The purchase of multimodal imaging equipment and the deployment of AI systems require significant upfront investment (147). Healthcare institutions and health policy decision-makers need to see clear evidence that this investment will bring long-term economic benefits through early intervention, reduced costs for treating complications, and optimized allocation of medical resources (149). Finally, there are challenges related to physician acceptance and ethical issues. Gaining the trust of clinicians is at the core of AI application (150). This depends not only on XAI technology but also requires large-scale prospective clinical trials to prove its effectiveness and safety in the real world. In addition, ethical and legal issues, such as potential algorithmic bias and the attribution of responsibility for medical errors, also need to be fully discussed and resolved before the technology is widely promoted (151, 152).

## 6 Conclusion and outlook

Multimodal AI is beginning a profound transformation, with the potential to reshape the management of DR by promoting a fundamental paradigm shift from standardized grading based on static, single-source information to personalized decision-making that integrates multi-source, dynamic data. By fusing ophthalmic imaging, systemic data, and emerging "omics" information, multimodal AI can construct an unprecedented "panoramic" digital model for each patient. This has the potential not only to significantly improve the accuracy of early warning, precise staging, and dynamic progression prediction for DR but also to provide a solid data foundation for formulating personalized treatment strategies.

Although the application of multimodal AI in the DR field is still in an exploratory stage, facing multiple challenges such as the scarcity of high-quality longitudinal datasets, the need to validate model generalization, and unclear clinical translation pathways, this has not diminished its great potential as a core driving force for future precision ophthalmology. Future research should focus on constructing large-scale, standardized multimodal public databases, developing more efficient and explainable fusion algorithms, and conducting prospective clinical validation studies

to accelerate the translation of technology from the lab to the clinic.

In summary, multimodal AI is not merely an improvement on existing diagnostic tools but a transformative force that can run through the entire process of DR management. It provides a novel perspective and powerful tools for conquering the challenge of blindness caused by this complex disease.

## Author contributions

CL: Methodology, Conceptualization, Writing – original draft. YD: Investigation, Writing – review & editing. HW: Validation, Writing – review & editing. JD: Conceptualization, Writing – review & editing, Methodology.

## Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was funded by the National Natural Science Foundation of China (grant number 82474394), the Noncommunicable Chronic Diseases-National Science and Technology Major Project (grant number 2023ZD0509300), and the CACMS Innovation Fund (grant number CI2023C007LH).

## References

- Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Res Clin Pract.* (2022) 183:109119. doi: 10.1016/j.diabres.2021.109119
- The Lancet. Diabetes: a defining disease of the 21st century. *Lancet.* (2023) 401:2087. doi: 10.1016/S0140-6736(23)01296-5
- Teo ZL, Tham YC, Yu M, Chee ML, Rim TH, Cheung N, et al. Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis. *Ophthalmology.* (2021) 128:1580–91. doi: 10.1016/j.ophtha.2021.04.027
- Pedersen HE, Sandvik CH, Subhi Y, Grauslund J, Pedersen FN. Relationship between diabetic retinopathy and systemic neurodegenerative diseases: a systematic review and meta-analysis. *Ophthalmol Retina.* (2022) 6:139–52. doi: 10.1016/j.oret.2021.07.002
- Kropp M, Golubnitschaja O, Mazurkova A, Koklesova L, Sargheini N, Vo TKS, et al. Diabetic retinopathy as the leading cause of blindness and early predictor of cascading complications-risks and mitigation. *EPMA J.* (2023) 14:21–42. doi: 10.1007/s13167-023-00314-8
- Tan TE, Wong TY. Diabetic retinopathy: looking forward to 2030. *Front Endocrinol.* (2022) 13:1077669. doi: 10.3389/fendo.2022.1077669
- Zhang Q, Zhang P, Chen N, Zhu Z, Li W, Wang Q. Trends and hotspots in the field of diabetic retinopathy imaging research from 2000–2023. *Front Med.* (2024) 11:1481088. doi: 10.3389/fmed.2024.1481088
- Chen Y, Song F, Zhao Z, Wang Y, To E, Liu Y, et al. Acceptability, applicability, and cost-utility of artificial-intelligence-powered low-cost portable fundus camera for diabetic retinopathy screening in primary health care settings. *Diabetes Res Clin Pract.* (2025) 223:112161. doi: 10.1016/j.diabres.2025.112161
- Gautam A, Shanker R. Diabetic retinopathy detection from fundus images: a wide survey from grading to segmentation of lesions. *Comput Biol Med.* (2025) 196(Pt B):110715. doi: 10.1016/j.compbiomed.2025.110715
- Baltrusaitis T, Ahuja C, Morency LP. Multimodal machine learning: a survey and taxonomy. *IEEE Trans Pattern Anal Mach Intell.* (2019) 41:423–43. doi: 10.1109/TPAMI.2018.2798607
- Sivaprasad S, Sen S, Cunha-Vaz J. Perspectives of diabetic retinopathy-challenges and opportunities. *Eye.* (2023) 37:2183–91. doi: 10.1038/s41433-022-02335-5
- Zhang Z, Deng C, Paulus YM. Advances in structural and functional retinal imaging and biomarkers for early detection of diabetic retinopathy. *Biomedicines.* (2024) 12:1405. doi: 10.3390/biomedicines12071405
- Tsiknakis N, Theodoropoulos D, Manikis G, Ktistakis E, Boutsora O, Berto A, et al. Deep learning for diabetic retinopathy detection and classification based on fundus images: a review. *Comput Biol Med.* (2021) 135:104599. doi: 10.1016/j.compbiomed.2021.104599
- Mohan NJ, Murugan R, Goel T, Roy P. Fast and robust exudate detection in retinal fundus images using extreme learning machine autoencoders and modified KAZE features. *J Digit Imaging.* (2022) 35:496–513. doi: 10.1007/s10278-022-00587-x
- Marcus DM, Silva PS, Liu D, Aiello LP, Antoszyk A, Elman M, et al. Association of predominantly peripheral lesions on ultra-widefield imaging and the risk of diabetic retinopathy worsening over time. *JAMA Ophthalmol.* (2022) 140:946–54. doi: 10.1001/jamaophthalmol.2022.3131
- Silva PS, Cavallerano JD, Haddad NM, Kwak H, Dyer KH, Omar AF, et al. Peripheral lesions identified on ultrawide field imaging predict increased risk of diabetic retinopathy progression over 4 years. *Ophthalmology.* (2015) 122:949–56. doi: 10.1016/j.ophtha.2015.01.008
- Patel SN, Shi A, Wibbelsman TD, Klufas MA. Ultra-widefield retinal imaging: an update on recent advances. *Ther Adv Ophthalmol.* (2020) 12:2515841419899495. doi: 10.1177/2515841419899495
- Aiello LP, Odia I, Glassman AR, Melia M, Jampol LM, Bressler NM, et al. Comparison of early treatment diabetic retinopathy study standard 7-field imaging with ultrawide-field imaging for determining severity of diabetic retinopathy. *JAMA Ophthalmol.* (2019) 137:65–73. doi: 10.1001/jamaophthalmol.2018.4982
- Wang M, Zhou X, Liu DN, Chen J, Zheng Z, Ling S. Development and validation of a predictive risk model based on retinal geometry for an early assessment of diabetic retinopathy. *Front Endocrinol.* (2022) 13:1033611. doi: 10.3389/fendo.2022.1033611
- Popovic N, Lipovac M, Radunovic M, Ugarte J, Isusquiza E, Beristain A, et al. Fractal characterization of retinal microvascular network morphology during diabetic retinopathy progression. *Microcirculation.* (2019) 26:e12531. doi: 10.1111/micc.12531
- Sabanayagam C, Lye WK, Klein R, Klein BE, Cotch MF, Wang JJ, et al. Retinal microvascular calibre and risk of diabetes mellitus: a systematic review and participant-level meta-analysis. *Diabetologia.* (2015) 58:2476–85. doi: 10.1007/s00125-015-3717-2

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

22. Tsai AS, Wong TY, Lavanya R, Zhang R, Hamzah H, Tai ES, et al. Differential association of retinal arteriolar and venular caliber with diabetes and retinopathy. *Diabetes Res Clin Pract.* (2011) 94:291–8. doi: 10.1016/j.diabres.2011.07.032
23. Song Y, Zhou Z, Liu H, Du R, Zhou Y, Zhu S, et al. Tortuosity of branch retinal artery is more associated with the genesis and progress of diabetic retinopathy. *Front Endocrinol.* (2022) 13:972339. doi: 10.3389/fendo.2022.972339
24. Forster RB, Garcia ES, Sluiman AJ, Grecian SM, McLachlan S, MacGillivray TJ, et al. Retinal venular tortuosity and fractal dimension predict incident retinopathy in adults with type 2 diabetes: the Edinburgh Type 2 Diabetes Study. *Diabetologia.* (2021) 64:1103–12. doi: 10.1007/s00125-021-05388-5
25. Classification of diabetic retinopathy from fluorescein angiograms. ETDRS report number 11. Early Treatment Diabetic Retinopathy Study Research Group. *Ophthalmology.* (1991) 98(5 Suppl.):807–22. doi: 10.1016/S0161-6420(13)38013-0
26. Fang M, Fan W, Shi Y, Ip MS, Wykoff CC, Wang K, et al. Classification of regions of nonperfusion on ultra-widefield fluorescein angiography in patients with diabetic macular edema. *Am J Ophthalmol.* (2019) 206:74–81. doi: 10.1016/j.ajo.2019.03.030
27. Ehlers JP, Jiang AC, Boss JD, Hu M, Figueiredo N, Babiuch A, et al. Quantitative ultra-widefield angiography and diabetic retinopathy severity: an assessment of parretinal leakage index, ischemic index and microaneurysm count. *Ophthalmology.* (2019) 126:1527–32. doi: 10.1016/j.ophtha.2019.05.034
28. Photocoagulation for diabetic macular edema. Early Treatment Diabetic Retinopathy Study report number 1. Early Treatment Diabetic Retinopathy Study research group. *Arch Ophthalmol.* (1985) 103:1796–806. doi: 10.1001/archophth.1985.01050120030015
29. Antaki F, Coussa RG, Mikhail M, Archambault C, Lederer DE. The prognostic value of peripheral retinal nonperfusion in diabetic retinopathy using ultra-widefield fluorescein angiography. *Graefes Arch Clin Exp Ophthalmol.* (2020) 258:2681–90. doi: 10.1007/s00417-020-04847-w
30. Tan GS, Cheung N, Simó R, Cheung GC, Wong TY. Diabetic macular oedema. *Lancet Diabetes Endocrinol.* (2017) 5:143–55. doi: 10.1016/S2213-8587(16)30052-3
31. Virgili G, Menchini F, Casazza G, Hogg R, Das RR, Wang X, et al. Optical coherence tomography (OCT) for detection of macular oedema in patients with diabetic retinopathy. *Cochrane Database Syst Rev.* (2015) 1:Cd008081. doi: 10.1002/14651858.CD008081.pub3
32. Cunha-Vaz J, Santos T, Ribeiro L, Alves D, Marques I, Goldberg M. OCT-leakage: a new method to identify and locate abnormal fluid accumulation in diabetic retinal edema. *Invest Ophthalmol Vis Sci.* (2016) 57:6776–83. doi: 10.1167/iov.16-19999
33. Gong Y, Wang M, Li Q, Shao Y, Li X. Evaluating the effect of vitreomacular interface abnormalities on anti-vascular endothelial growth factor treatment outcomes in diabetic macular edema by optical coherence tomography: a systematic review and meta-analysis. *Photodiagnosis Photodyn Ther.* (2023) 42:103555. doi: 10.1016/j.pdpdt.2023.103555
34. Sun JK, Radwan SH, Soliman AZ, Lammer J, Lin MM, Prager SG, et al. Neural retinal disorganization as a robust marker of visual acuity in current and resolved diabetic macular edema. *Diabetes.* (2015) 64:2560–70. doi: 10.2337/db14-0782
35. Zafar S, Sachdeva M, Frankfort BJ, Channa R. Retinal neurodegeneration as an early manifestation of diabetic eye disease and potential neuroprotective therapies. *Curr Diab Rep.* (2019) 19:17. doi: 10.1007/s11892-019-1134-5
36. Wanek J, Blair NP, Chau FY, Lim JI, Leiderman YI, Shahidi M. Alterations in retinal layer thickness and reflectance at different stages of diabetic retinopathy by en face optical coherence tomography. *Invest Ophthalmol Vis Sci.* (2016) 57:OCT341–7. doi: 10.1167/iov.15-18715
37. Vujosevic S, Midea E. Retinal layers changes in human preclinical and early clinical diabetic retinopathy support early retinal neuronal and Müller cells alterations. *J Diabetes Res.* (2013) 2013:905058. doi: 10.1155/2013/905058
38. Lim HB, Shin YI, Lee MW, Koo H, Lee WH, Kim JY. Ganglion cell - inner plexiform layer damage in diabetic patients: 3-year prospective, longitudinal, observational study. *Sci Rep.* (2020) 10:1470. doi: 10.1038/s41598-020-58465-x
39. Sohn EH, van Dijk HW, Jiao C, Kok PH, Jeong W, Demirkaya N, et al. Retinal neurodegeneration may precede microvascular changes characteristic of diabetic retinopathy in diabetes mellitus. *Proc Natl Acad Sci USA.* (2016) 113:E2655–64. doi: 10.1073/pnas.1522014113
40. Zhang F, Du Z, Zhang X, Wang Y, Chen Y, Wu G, et al. Alterations of outer retinal reflectivity in diabetic patients without clinically detectable retinopathy. *Graefes Arch Clin Exp Ophthalmol.* (2024) 262:61–72. doi: 10.1007/s00417-023-06238-3
41. Cetin EN, Parca O, Akkaya HS, Pekel G. Association of inner retinal reflectivity with qualitative and quantitative changes in retinal layers over time in diabetic eyes without retinopathy. *Eye.* (2022) 36:1253–60. doi: 10.1038/s41433-021-01607-w
42. Tey KY, Teo K, Tan ACS, Devarajan K, Tan B, Tan J, et al. Optical coherence tomography angiography in diabetic retinopathy: a review of current applications. *Eye Vis.* (2019) 6:37. doi: 10.1186/s40662-019-0160-3
43. Wijesingha N, Tsai WS, Keskin AM, Holmes C, Kazantzis D, Chandak S, et al. Optical coherence tomography angiography as a diagnostic tool for diabetic retinopathy. *Diagnostics.* (2024) 14:326. doi: 10.3390/diagnostics14030326
44. Kawai K, Murakami T, Mori Y, Ishihara K, Dodo Y, Terada N, et al. Clinically significant nonperfusion areas on widefield OCT angiography in diabetic retinopathy. *Ophthalmol Sci.* (2023) 3:100241. doi: 10.1016/j.xops.2022.100241
45. Sun Z, Tang F, Wong R, Lok J, Szeto SKH, Chan JCK, et al. OCT angiography metrics predict progression of diabetic retinopathy and development of diabetic macular edema: a prospective study. *Ophthalmology.* (2019) 126:1675–84. doi: 10.1016/j.ophtha.2019.06.016
46. Custo Greig E, Brigell M, Cao F, Levine ES, Peters K, Moulton EM, et al. Macular and peripapillary optical coherence tomography angiography metrics predict progression in diabetic retinopathy: a sub-analysis of TIME-2b study data. *Am J Ophthalmol.* (2020) 219:66–76. doi: 10.1016/j.ajo.2020.06.009
47. Wang H, Liu X, Hu X, Xin H, Bao H, Yang S. Retinal and choroidal microvascular characterization and density changes in different stages of diabetic retinopathy eyes. *Front Med.* (2023) 10:1186098. doi: 10.3389/fmed.2023.1186098
48. Yuan M, Wang W, Kang S, Li Y, Li W, Gong X, et al. Peripapillary microvasculature predicts the incidence and development of diabetic retinopathy: an SS-OCTA study. *Am J Ophthalmol.* (2022) 243:19–27. doi: 10.1016/j.ajo.2022.07.001
49. Song KH, Jeong JS, Kim MK, Kwon HS, Baek KH, Ko SH, et al. Discordance in risk factors for the progression of diabetic retinopathy and diabetic nephropathy in patients with type 2 diabetes mellitus. *J Diabetes Investig.* (2019) 10:745–52. doi: 10.1111/jdi.12953
50. Liu L, Quang ND, Banu R, Kumar H, Tham YC, Cheng CY, et al. Hypertension, blood pressure control and diabetic retinopathy in a large population-based study. *PLoS ONE.* (2020) 15:e0229665. doi: 10.1371/journal.pone.0229665
51. Ivanescu A, Popescu S, Ivanescu R, Potra M, Timar R. Predictors of diabetic retinopathy in type 2 diabetes: a cross-sectional study. *Biomedicines.* (2024) 12:1889. doi: 10.3390/biomedicines12081889
52. Kaur A, Kumar R, Sharma A. Diabetic retinopathy leading to blindness—a review. *Curr Diabetes Rev.* (2024) 20:e240124225997. doi: 10.2174/0115733998274599231109034741
53. Cho A, Park HC, Lee YK, Shin YJ, Bae SH, Kim H. Progression of diabetic retinopathy and declining renal function in patients with type 2 diabetes. *J Diabetes Res.* (2020) 2020:8784139. doi: 10.1155/2020/8784139
54. Tsui CK, Hu A, Li Y, Huang W, Wang W, Liu K, et al. Prevalence, incidence, and risk factors of diabetic retinopathy and macular edema in patients with early and late-onset type 2 diabetes mellitus. *J Diabetes Investig.* (2025) 16:1254–62. doi: 10.1111/jdi.70027
55. Zhang G, Chen W, Chen H, Lin J, Cen LP, Xie P, et al. Risk factors for diabetic retinopathy, diabetic macular edema, and sight-threatening diabetic retinopathy. *Asia Pac J Ophthalmol.* (2024) 13:100067. doi: 10.1016/j.apjo.2024.100067
56. Cai Y, Qiu W, Ma X, Yang Y, Tang T, Dong Y, et al. Association between renal function and diabetic retinopathy: a mediation analysis of geriatric nutritional risk index. *Diabetol Metab Syndr.* (2025) 17:95. doi: 10.1186/s13098-025-01658-z
57. Du K, Luo W. Association between blood urea nitrogen levels and diabetic retinopathy in diabetic adults in the United States (NHANES 2005-2018). *Front Endocrinol.* (2024) 15:1403456. doi: 10.3389/fendo.2024.1403456
58. Dongling N, Ziwei K, Juanling S, Li Z, Chang W, Ting L, et al. Universal nomogram for predicting referable diabetic retinopathy: a validated model for community and ophthalmic outpatient populations using easily accessible indicators. *Front Endocrinol.* (2025) 16:1557166. doi: 10.3389/fendo.2025.1557166
59. Lian XN, Zhu MM. Factors related to type 2 diabetic retinopathy and their clinical application value. *Front Endocrinol.* (2024) 15:1484197. doi: 10.3389/fendo.2024.1484197
60. Tabakoglu NT, Celik M. Investigation of the systemic immune inflammation (SII) index as an indicator of morbidity and mortality in type 2 diabetic retinopathy patients in a 4-year follow-up period. *Medicina.* (2024) 60:855. doi: 10.20944/preprints202405.0211.v1
61. Wang S, Pan X, Jia B, Chen S. Exploring the correlation between the systemic immune inflammation index (SII), systemic inflammatory response index (SIRI), and type 2 diabetic retinopathy. *Diabetes Metab Syndr Obes.* (2023) 16:3827–36. doi: 10.2147/DMSO.S437580
62. Wu Q, Zhao B, Dongye S, Sun L, An B, Xu Q. Systemic inflammatory markers and neurovascular changes in the retina and choroid of diabetic patients without retinopathy: insights from wide-field SS-OCTA. *Front Med.* (2025) 12:1566047. doi: 10.3389/fmed.2025.1566047
63. Deng R, Zhu S, Fan B, Chen X, Lv H, Dai Y. Exploring the correlations between six serological inflammatory markers and different stages of type 2 diabetic retinopathy. *Sci Rep.* (2025) 15:1567. doi: 10.1038/s41598-025-85164-2
64. Imamura M, Takahashi A, Matsunami M, Horikoshi M, Iwata M, Araki SI, et al. Genome-wide association studies identify two novel loci conferring susceptibility to diabetic retinopathy in Japanese patients with type 2 diabetes. *Hum Mol Genet.* (2021) 30:716–26. doi: 10.1093/hmg/ddab044
65. Dieter C, Lemos NE, de Faria Corrêa NR, Assmann TS, Pellenz FM, Canani LH, et al. Polymorphisms in TIE2 and ANGPT-1 genes are associated with protection against diabetic retinopathy in a Brazilian population. *Arch Endocrinol Metab.* (2023) 67:e000624. doi: 10.20945/2359-399700000624

66. Iwasaki H, Shichiri M. Protein arginine N-methyltransferase 1 gene polymorphism is associated with proliferative diabetic retinopathy in a Japanese population. *Acta Diabetol.* (2022) 59:319–27. doi: 10.1007/s00592-021-01808-5
67. Jafarzadeh F, Javanbakht A, Bakhtar N, Dalvand A, Shabani M, Mehrabinejad MM. Association between diabetic retinopathy and polymorphisms of cytokine genes: a systematic review and meta-analysis. *Int Ophthalmol.* (2022) 42:349–61. doi: 10.1007/s10792-021-02011-9
68. Hu L, Gong C, Chen X, Zhou H, Yan J, Hong W. Associations between vascular endothelial growth factor gene polymorphisms and different types of diabetic retinopathy susceptibility: a systematic review and meta-analysis. *J Diabetes Res.* (2021) 2021:7059139. doi: 10.1155/2021/7059139
69. Sienkiewicz-Szapka E, Fiedorowicz E, Król-Grzymała A, Kordulewska N, Rozmus D, Cieślińska A, et al. The role of genetic polymorphisms in diabetic retinopathy: narrative review. *Int J Mol Sci.* (2023) 24:15865. doi: 10.3390/ijms242115865
70. Zhou X, Hou G, Wang X, Peng Z, Yin X, Yang J, et al. Metabolomic studies reveal and validate potential biomarkers of diabetic retinopathy in two Chinese datasets with type 2 diabetes: a cross-sectional study. *Cardiovasc Diabetol.* (2024) 23:439. doi: 10.1186/s12933-024-02535-1
71. Fernandes Silva L, Hokkanen J, Vangipurapu J, Oravilaiti A, Laakso M. Metabolites as risk factors for diabetic retinopathy in patients with type 2 diabetes: a 12-year follow-up study. *J Clin Endocrinol Metab.* (2023) 109:100–6. doi: 10.1210/clinem/dgad452
72. Wen X, Ng TK, Zhang G, Chen H, Wu Z, Liu Q, et al. Tear lactate improves the evaluation of proliferative diabetic retinopathy in type-2 diabetes patients. *Mol Biomed.* (2025) 6:52. doi: 10.1186/s43556-025-00297-0
73. Wang H, Fang J, Chen F, Sun Q, Xu X, Lin SH, et al. Metabolomic profile of diabetic retinopathy: a GC-TOFMS-based approach using vitreous and aqueous humor. *Acta Diabetol.* (2020) 57:41–51. doi: 10.1007/s00592-019-01363-0
74. García-Ramírez M, Canals F, Hernández C, Colomé N, Ferrer C, Carrasco E, et al. Proteomic analysis of human vitreous fluid by fluorescence-based difference gel electrophoresis (DIGE): a new strategy for identifying potential candidates in the pathogenesis of proliferative diabetic retinopathy. *Diabetologia.* (2007) 50:1294–303. doi: 10.1007/s00125-007-0627-y
75. Chiang SY, Tsai ML, Wang CY, Chen A, Chou YC, Hsia CW, et al. Proteomic analysis and identification of aqueous humor proteins with a pathophysiological role in diabetic retinopathy. *J Proteomics.* (2012) 75:2950–9. doi: 10.1016/j.jpro.2011.12.006
76. Wolf J, Rasmussen DK, Sun YJ, Vu JT, Wang E, Espinosa C, et al. Liquid-biopsy proteomics combined with AI identifies cellular drivers of eye aging and disease *in vivo*. *Cell.* (2023) 186:4868–84.e12. doi: 10.1016/j.cell.2023.09.012
77. Wang S, He X, Jian Z, Li J, Xu C, Chen Y, et al. Advances and prospects of multi-modal ophthalmic artificial intelligence based on deep learning: a review. *Eye Vis.* (2024) 11:38. doi: 10.1186/s40662-024-00405-1
78. Chang X, Cai L, Wang J, Dong H, Han J, Wang C. Sparse-view photoacoustic reconstruction method for diabetic retinopathy using feature fusion network. *J Biophotonics.* (2024) 17:e202400287. doi: 10.1002/jbio.202400287
79. Khalid M, Sajid MZ, Youssef A, Khan NA, Hamid MF, Abbas F, et al. An automated system for multi-eye disease classification using feature fusion with deep learning models and fluorescence imaging for enhanced interpretability. *Diagnostics.* (2024) 14:2679. doi: 10.3390/diagnostics14232679
80. Ejaz S, Zia HU, Majeed F, Shafique U, Altamiranda SC, Lipari V, et al. Fundus image classification using feature concatenation for early diagnosis of retinal disease. *Digit Health.* (2025) 11:20552076251328120. doi: 10.1177/20552076251328120
81. Hervella AS, Rouco J, Novo J, Ortega M. Multimodal image encoding pre-training for diabetic retinopathy grading. *Comput Biol Med.* (2022) 143:105302. doi: 10.1016/j.compbiomed.2022.105302
82. Huang SC, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *NPJ Digit Med.* (2020) 3:136. doi: 10.1038/s41746-020-00341-z
83. Qummar S, Khan FG, Shah S, Khan A, Shamshirband S, Rehman ZU, et al. A deep learning ensemble approach for diabetic retinopathy detection. *IEEE Access.* (2019) 7:150530–9. doi: 10.1109/ACCESS.2019.2947484
84. Sekar P, Suba Raja KS, Krishnaraj R. DRCNN-Lesion Proxy: a hybrid CNN with lesion-inspired feature simulation for diabetic retinopathy severity classification. *Sci Rep.* (2025) 15:37954. doi: 10.1038/s41598-025-21337-3
85. Ebrahimi B, Le D, Abtahi M, Dadzie AK, Lim JJ, Chan RVP, et al. Optimizing the OCTA layer fusion option for deep learning classification of diabetic retinopathy. *Biomed Opt Express.* (2023) 14:4713–24. doi: 10.1364/BOE.495999
86. Li Y, El Habib Daho M, Conze PH, Zeglache R, Le Boité H, Bonnin S, et al. Hybrid fusion of high-resolution and ultra-widefield OCTA acquisitions for the automatic diagnosis of diabetic retinopathy. *Diagnostics.* (2023) 13:2770. doi: 10.3390/diagnostics13172770
87. Tseng VS, Chen CL, Liang CM, Tai MC, Liu JT, Wu PY, et al. Leveraging multimodal deep learning architecture with retinal lesion information to detect diabetic retinopathy. *Transl Vis Sci Technol.* (2020) 9:41. doi: 10.1167/tvst.9.2.41
88. Ai Z, Huang X, Fan Y, Feng J, Zeng F, Lu Y, et al. Detection algorithm of diabetic retinopathy based on deep ensemble learning and attention mechanism. *Front Neuroinform.* (2021) 15:778552. doi: 10.3389/fninf.2021.778552
89. Zang F, Ma H. CRA-Net: transformer guided category-relation attention network for diabetic retinopathy grading. *Comput Biol Med.* (2024) 170:107993. doi: 10.1016/j.compbiomed.2024.107993
90. Ikram A, Imran A. ResViT FusionNet Model: an explainable AI-driven approach for automated grading of diabetic retinopathy in retinal images. *Comput Biol Med.* (2025) 186:109656. doi: 10.1016/j.compbiomed.2025.109656
91. Rossi D, Citarella AA, De Marco F, Di Biasi L, Tortora G. Comparative analysis of diabetes diagnosis: WE-LSTM networks and WizardLM-powered DiabeTalk chatbot. In: *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. New York, NY: IEEE (2024). p. 6859–66. doi: 10.1109/BIBM62325.2024.10821742
92. Watson DS, Krutzinna J, Bruce IN, Griffiths CE, McInnes IB, Barnes MR, et al. Clinical applications of machine learning algorithms: beyond the black box. *BMJ.* (2019) 364:l886. doi: 10.1136/bmj.l886
93. Quellec G, Al Hajj H, Lamard M, Conze PH, Massin P, Cochener B. ExplAI: explanatory artificial intelligence for diabetic retinopathy diagnosis. *Med Image Anal.* (2021) 72:102118. doi: 10.1016/j.media.2021.102118
94. Sharma N, Lalwani P. A multi model deep net with an explainable AI based framework for diabetic retinopathy segmentation and classification. *Sci Rep.* (2025) 15:8777. doi: 10.1038/s41598-025-93376-9
95. Rautaray J, Ali ABM, Kandpal M, Mishra P, Rashid RF, Alimova F, et al. Leveraging FastViT based knowledge distillation with EfficientNet-B0 for diabetic retinopathy severity classification. *SLAS Technol.* (2025) 33:100325. doi: 10.1016/j.slant.2025.100325
96. Li H, Dong X, Shen W, Ge F, Li H. Resampling-based cost loss attention network for explainable imbalanced diabetic retinopathy grading. *Comput Biol Med.* (2022) 149:105970. doi: 10.1016/j.compbiomed.2022.105970
97. Delavari P, Ozturan G, Navajas EV, Yilmaz O, Oruc I. AI-assisted identification of sex-specific patterns in diabetic retinopathy using retinal fundus images. *PLoS ONE.* (2025) 20:e0327305. doi: 10.1371/journal.pone.0327305
98. Shin J. Feasibility of local interpretable model-agnostic explanations (LIME) algorithm as an effective and interpretable feature selection method: comparative fNIRS study. *Biomed Eng Lett.* (2023) 13:689–703. doi: 10.1007/s13534-023-00291-x
99. Nohara Y, Matsumoto K, Soejima H, Nakashima N. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Comput Methods Programs Biomed.* (2022) 214:106584. doi: 10.1016/j.cmpb.2021.106584
100. Zong GW, Wang WY, Zheng J, Zhang W, Luo WM, Fang ZZ, et al. A metabolism-based interpretable machine learning prediction model for diabetic retinopathy risk: a cross-sectional study in Chinese patients with type 2 diabetes. *J Diabetes Res.* (2023) 2023:3990035. doi: 10.1155/2023/3990035
101. Yagin FH, Colak C, Algarni A, Gormez Y, Guldogan E, Ardigò LP. Hybrid explainable artificial intelligence models for targeted metabolomics analysis of diabetic retinopathy. *Diagnostics.* (2024) 14:1364. doi: 10.3390/diagnostics14131364
102. Gui Y, Gui S, Wang X, Li Y, Xu Y, Zhang J. Exploring the relationship between heavy metals and diabetic retinopathy: a machine learning modeling approach. *Sci Rep.* (2024) 14:13049. doi: 10.1038/s41598-024-63916-w
103. Xu X, Zhang M, Huang S, Li X, Kui X, Liu J. The application of artificial intelligence in diabetic retinopathy: progress and prospects. *Front Cell Dev Biol.* (2024) 12:1473176. doi: 10.3389/fcell.2024.1473176
104. Liew G, Michaelides M, Bunce C. A comparison of the causes of blindness certifications in England and Wales in working age adults (16–64 years), 1999–2000 with 2009–2010. *BMJ Open.* (2014) 4:e004015. doi: 10.1136/bmjopen-2013-004015
105. Hsu MY, Chiou JY, Liu JT, Lee CM, Lee YW, Chou CC, et al. Deep learning for automated diabetic retinopathy screening fused with heterogeneous data from EHRs can lead to earlier referral decisions. *Transl Vis Sci Technol.* (2021) 10:18. doi: 10.1167/tvst.10.9.18
106. Messica S, Cohen S, Hadad A, Gordon M, Katz O, Presil D, et al. Temporal integrative machine learning for early detection of diabetic retinopathy using fundus imaging and electronic health records. *IEEE J Biomed Health Inform.* (2025). doi: 10.1109/JBHI.2025.3578197. [Epub ahead of print].
107. Liu Z, Han X, Gao L, Chen S, Huang W, Li P, et al. Cost-effectiveness of incorporating self-imaging optical coherence tomography into fundus photography-based diabetic retinopathy screening. *NPJ Digit Med.* (2024) 7:225. doi: 10.1038/s41746-024-01222-5
108. Zedadra A, Salah-Salah MY, Zedadra O, Guerrieri A. Multi-modal AI for multi-label retinal disease prediction using OCT and fundus images: a hybrid approach. *Sensors.* (2025) 25:4492. doi: 10.3390/s25144492
109. Wong TY, Sun J, Kawasaki R, Ruamviboonsuk P, Gupta N, Lansingh VC, et al. Guidelines on diabetic eye care: the international council of ophthalmology

- recommendations for screening, follow-up, referral, and treatment based on resource settings. *Ophthalmology*. (2018) 125:1608–22. doi: 10.1016/j.ophtha.2018.04.007
110. Pan H, Miao J, Yu J, Li J, Wang X, Feng J. Multi-modal classification of retinal disease based on convolutional neural network. *Biomed Phys Eng Express*. (2025) 11:45030. doi: 10.1088/2057-1976/adeb92
111. Hajeb Mohammad Alipour S, Rabbani H, Akhlaghi MR. Diabetic retinopathy grading by digital curvelet transform. *Comput Math Methods Med*. (2012) 2012:761901. doi: 10.1155/2012/761901
112. Sandhu HS, Elmogy M, Taher Sharafeldeen A, Elsharkawy M, El-Adawy N, Eltanboly A, et al. Automated diagnosis of diabetic retinopathy using clinical biomarkers, optical coherence tomography, and optical coherence tomography angiography. *Am J Ophthalmol*. (2020) 216:201–6. doi: 10.1016/j.ajo.2020.01.016
113. Torok Z, Peto T, Csoz E, Tukacs E, Molnar AM, Berta A, et al. Combined methods for diabetic retinopathy screening, using retina photographs and tear fluid proteomics biomarkers. *J Diabetes Res*. (2015) 2015:623619. doi: 10.1155/2015/623619
114. Li X, Jia M, Islam MT, Yu L, Xing L. Self-supervised feature learning via exploiting multi-modal data for retinal disease diagnosis. *IEEE Trans Med Imaging*. (2020) 39:4023–33. doi: 10.1109/TMI.2020.3008871
115. He X, Deng Y, Fang L, Peng Q. Multi-modal retinal image classification with modality-specific attention network. *IEEE Trans Med Imaging*. (2021) 40:1591–602. doi: 10.1109/TMI.2021.3059956
116. Hua CH, Kim K, Huynh-The T, You JI, Yu SY, Le-Tien T, et al. Convolutional network with twofold feature augmentation for diabetic retinopathy recognition from multi-modal images. *IEEE J Biomed Health Inform*. (2021) 25:2686–97. doi: 10.1109/JBHI.2020.3041848
117. Song J, Zheng Y, Wang J, Zakir Ullah M, Jiao W. Multicolor image classification using the multimodal information bottleneck network (MMIB-Net) for detecting diabetic retinopathy. *Opt Express*. (2021) 29:22732–48. doi: 10.1364/OE.430508
118. Boudegga H, Elloumi Y, Akil M, Hedi Bedoui M, Kachouri R, Abdallah AB. Fast and efficient retinal blood vessel segmentation method based on deep learning network. *Comput Med Imaging Graph*. (2021) 90:101902. doi: 10.1016/j.compmedimag.2021.101902
119. Wang S, Yu X, Wu H, Wang Y, Wu C. CMFNet: a cross-dimensional modal fusion network for accurate vessel segmentation based on OCTA data. *Med Biol Eng Comput*. (2025) 63:1161–76. doi: 10.1007/s11517-024-03256-z
120. Xie Q, Li X, Li Y, Lu J, Ma S, Zhao Y, et al. A multi-modal multi-branch framework for retinal vessel segmentation using ultra-widefield fundus photographs. *Front Cell Dev Biol*. (2024) 12:1532228. doi: 10.3389/fcell.2024.1532228
121. Rodrigues EO, Conci A, Liatsis P. ELEMENT: multi-modal retinal vessel segmentation based on a coupled region growing and machine learning approach. *IEEE J Biomed Health Inform*. (2020) 24:3507–19. doi: 10.1109/JBHI.2020.2999257
122. Bi Z, Li J, Liu Q, Fang Z. Deep learning-based optical coherence tomography and retinal images for detection of diabetic retinopathy: a systematic and meta analysis. *Front Endocrinol*. (2025) 16:1485311. doi: 10.3389/fendo.2025.1485311
123. Liang Y, Zhang X, Mei W, Li Y, Du Z, Wang Y, et al. Predicting vision-threatening diabetic retinopathy in patients with type 2 diabetes mellitus: systematic review, meta-analysis, and prospective validation study. *J Glob Health*. (2024) 14:4192. doi: 10.7189/jogh.14.04192
124. Jin K, Ye J. Artificial intelligence and deep learning in ophthalmology: current status and future perspectives. *Adv Ophthalmol Pract Res*. (2022) 2:100078. doi: 10.1016/j.aopr.2022.100078
125. Arcadu F, Benmansour F, Maunz A, Willis J, Haskova Z, Prunotto M. Deep learning algorithm predicts diabetic retinopathy progression in individual patients. *NPJ Digit Med*. (2019) 2:92. doi: 10.1038/s41746-019-0172-3
126. Dai L, Wu L, Li H, Cai C, Wu Q, Kong H, et al. A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nat Commun*. (2021) 12:3242. doi: 10.1038/s41467-021-23458-5
127. Han D, Gao J, Xu Z, Zhang C, Jiang B, Wei C, et al. Predictive model for proliferative diabetic retinopathy using single-cell transcriptomics. *Exp Eye Res*. (2025) 259:110536. doi: 10.1016/j.exer.2025.110536
128. Qiao H, Tang F, Zhou H, Cai Y, Guo K, Wang J, et al. Forecasting the diabetic retinopathy progression using generative adversarial networks. *Commun Med*. (2025) 5:368. doi: 10.1038/s43856-025-01092-2
129. Alsadoun L, Ali H, Mushtaq MM, Mushtaq M, Burhanuddin M, Anwar R, et al. Artificial intelligence (AI)-enhanced detection of diabetic retinopathy from fundus images: the current landscape and future directions. *Cureus*. (2024) 16:e67844. doi: 10.7759/cureus.67844
130. Kalra G, Wykoff C, Martin A, Srivastava SK, Reese J, Ehlers JP. Longitudinal quantitative ultrawidefield angiographic features in diabetic retinopathy treated with aflibercept from the intravitreal aflibercept as indicated by real-time objective imaging to achieve diabetic retinopathy improvement trial. *Ophthalmol Retina*. (2024) 8:116–25. doi: 10.1016/j.oret.2023.09.004
131. Pappuru RKR, Ribeiro L, Lobo C, Alves D, Cunha-Vaz J. Microaneurysm turnover is a predictor of diabetic retinopathy progression. *Br J Ophthalmol*. (2019) 103:222–6. doi: 10.1136/bjophthalmol-2018-311887
132. Pang Y, Luo C, Zhang Q, Zhang X, Liao N, Ji Y, et al. Multi-omics integration with machine learning identified early diabetic retinopathy, diabetic macula edema and anti-VEGF treatment response. *Transl Vis Sci Technol*. (2024) 13:23. doi: 10.1167/tvst.13.12.23
133. Low SWY, Vaidya T, Gadde SGK, Mochi TB, Kumar D, Kassem IS, et al. Decorin concentrations in aqueous humor of patients with diabetic retinopathy. *Life*. (2021) 11:1421. doi: 10.3390/life11121421
134. Jones CD, Greenwood RH, Misra A, Bachmann MO. Incidence and progression of diabetic retinopathy during 17 years of a population-based screening program in England. *Diabetes Care*. (2012) 35:592–6. doi: 10.2337/dc11-0943
135. Dai L, Sheng B, Chen T, Wu Q, Liu R, Cai C, et al. A deep learning system for predicting time to progression of diabetic retinopathy. *Nat Med*. (2024) 30:584–94. doi: 10.1038/s41591-023-02742-5
136. Nguyen HV, Tan GS, Tapp RJ, Mital S, Ting DS, Wong HT, et al. Cost-effectiveness of a national telemedicine diabetic retinopathy screening program in Singapore. *Ophthalmology*. (2016) 123:2571–80. doi: 10.1016/j.ophtha.2016.08.021
137. Irodi A, Zhu Z, Grzybowski A, Wu Y, Cheung CY, Li H, et al. The evolution of diabetic retinopathy screening. *Eye*. (2025) 39:1040–6. doi: 10.1038/s41433-025-03633-4
138. Quinn TP, Senadeera M, Jacobs S, Coghlan S, Le V. Trust and medical AI: the challenges we face and the expertise needed to overcome them. *J Am Med Inform Assoc*. (2021) 28:890–4. doi: 10.1093/jamia/ocaa268
139. Grzybowski A, Jin K, Zhou J, Pan X, Wang M, Ye J, et al. Retina fundus photograph-based artificial intelligence algorithms in medicine: a systematic review. *Ophthalmol Ther*. (2024) 13:2125–49. doi: 10.1007/s40123-024-00981-4
140. Granger BB, Shah BR. Blending quality improvement and research methods for implementation science, part I: design and data collection. *AACN Adv Crit Care*. (2015) 26:268–74. doi: 10.4037/NCL.0000000000000090
141. Wamat-Herresthal S, Schultze H, Shastry KL, Manamohan S, Mukherjee S, Garg V, et al. Swarm learning for decentralized and confidential clinical machine learning. *Nature*. (2021) 594:265–70. doi: 10.1038/s41586-021-03583-3
142. Farahat Z, Zrira N, Souissi N, Bennani Y, Bencherif S, Benamar S, et al. Diabetic retinopathy screening through artificial intelligence algorithms: a systematic review. *Surv Ophthalmol*. (2024) 69:707–21. doi: 10.1016/j.survophthal.2024.05.008
143. Guo M, Gong D, Yang W. In-depth analysis of research hotspots and emerging trends in AI for retinal diseases over the past decade. *Front Med*. (2024) 11:1489139. doi: 10.3389/fmed.2024.1489139
144. Yang Q, Bee YM, Lim CC, Sabanayagam C, Yim-Lui Cheung C, Wong TY, et al. Use of artificial intelligence with retinal imaging in screening for diabetes-associated complications: systematic review. *EclinicalMedicine*. (2025) 81:103089. doi: 10.1016/j.eclim.2025.103089
145. Poly TN, Islam MM, Walther BA, Lin MC, Jack Li YC. Artificial intelligence in diabetic retinopathy: bibliometric analysis. *Comput Methods Programs Biomed*. (2023) 231:107358. doi: 10.1016/j.cmpb.2023.107358
146. Poon AIF, Sung JY. Opening the black box of AI-medicine. *J Gastroenterol Hepatol*. (2021) 36:581–4. doi: 10.1111/jgh.15384
147. Nakayama LF, Zago Ribeiro L, Novaes F, Miyawaki IA, Miyawaki AE, de Oliveira JAE, et al. Artificial intelligence for telemedicine diabetic retinopathy screening: a review. *Ann Med*. (2023) 55:2258149. doi: 10.1080/07853890.2023.2258149
148. Hill DLG. AI in imaging: the regulatory landscape. *Br J Radiol*. (2024) 97:483–91. doi: 10.1093/bjr/tqae002
149. Cleland CR, Rwiza J, Evans JR, Gordon I, MacLeod D, Burton MJ, et al. Artificial intelligence for diabetic retinopathy in low-income and middle-income countries: a scoping review. *BMJ Open Diabetes Res Care*. (2023) 11:e003424. doi: 10.1136/bmjdr-2023-003424
150. Zhao M, Jiang Y. Great expectations and challenges of artificial intelligence in the screening of diabetic retinopathy. *Eye*. (2020) 34:418–9. doi: 10.1038/s41433-019-0629-2
151. Xu J, Xue K, Zhang K. Current status and future trends of clinical diagnoses via image-based deep learning. *Theranostics*. (2019) 9:7556–65. doi: 10.7150/thno.38065
152. Abdullah YI, Schuman JS, Shabsigh R, Caplan A, Al-Aswad LA. Ethics of artificial intelligence in medicine and ophthalmology. *Asia Pac J Ophthalmol*. (2021) 10:289–98. doi: 10.1097/APO.0000000000000397