# Multimodal Medicine in Glaucoma, Diabetic Retinopathy, and Age-related Macular Degeneration: Application and Prospect

Siyue Sun

Mathmatics, China University of Mining and Technology, Xu Zhou, JiangSu, China

SSY7559@outlook.com

**Abstract.** Glaucoma, diabetic retinopathy (DR), and age-related macular degeneration (AMD) are the leading causes of blindness worldwide, and single-modal techniques are insufficient to meet the demands of diagnosis and treatment. Therefore, this article reviews the research on multimodal medical technology in glaucoma, diabetic retinopathy, and age-related macular degeneration. In the diagnosis and treatment of early glaucoma, multimodal fusion can effectively improve the accuracy of diagnosis and grading, and further optimize related decisions. In the diagnosis and treatment of DR, the fusion architecture can improve accuracy, but the MMDA framework has scarce data. However, the DeepDR-LLM system can improve the efficiency of primary care. In the diagnosis and treatment of AMD, a dual-stream Convolutional Neural Network (CNN) can optimize classification effects, and anti-VEGF drugs have good therapeutic effects. DR has a protective effect on AMD. However, current technologies still have problems such as insufficient fusion. In the future, it is necessary to further optimize technical data, break through transformation, and promote the application of multimodal technology in ophthalmology.

**Keywords:** Glaucoma, DR, AMD, Multimodal.

## 1. Introduction

Eye health is the core of ensuring human visual function and quality of life. However, glaucoma, diabetic retinopathy (DR), and age-related macular degeneration (AMD) are the three leading causes of blindness globally, and their diagnosis and treatment processes face dual challenges of precision and timeliness [1, 2]. Glaucoma, due to the insidious nature of irreversible damage to the optic nerve, often results in missed diagnoses when relying solely on intraocular pressure monitoring or visual field tests for early detection. DR, associated with the progression of diabetes, presents complex retinal microvascular changes, and traditional retinal photography struggles to simultaneously quantify vascular leakage and the degree of nerve damage. AMD, with its delicate macular region and diverse lesion types, makes it difficult for a single imaging technique to comprehensively assess choroidal neovascularization activity and degenerative changes in the retinal outer layers [3, 4, 5].

With the development of medical technology towards multi-dimensional and cross-modal integration, multi-modal medical technology provides better solutions for the diagnosis and treatment of eye diseases [6, 7, 8]. This technology integrates multiple data sources such as optical coherence tomography (OCT), fundus fluorescein angiography (FFA), fundus autofluorescence (FAF), visual field examination, and artificial intelligence-assisted diagnostic algorithms to achieve a comprehensive analysis of the structure, function, and pathological features of the eye, effectively compensating for the deficiencies of single-modal technology in information coverage and diagnostic specificity [9, 10, 11].

This article reviews the three core ophthalmic diseases: glaucoma, DR, and AMD, summarizing the technological approaches and clinical value of data fusion from different modalities. It aims to provide a reference for promoting the standardized application and innovative development of multimodal medical technologies in the field of ocular treatment.

## 2. Research on Glaucoma Based on Multimodal Information

Glaucoma is a group of ophthalmic diseases characterized primarily by optic nerve damage and visual field defects, often associated with pathological elevation of intraocular pressure. If not treated promptly, it may lead to irreversible blindness.

As early as 2017, Omodaka et al. found from clinical experiments that age is a key factor in the pathogenesis of glaucoma, and machine learning can assist in early diagnosis. They trained a machine learning model using quantitative eye parameters such as age, allowing the deep learning model to integrate genomic data to identify genetic loci associated with vertical cup-and-dish ratio (VCDR) to classify the optic disc morphology of glaucoma patients. Experiments have found that age is positively correlated with VCDR, and VCDR enlargement is a key risk factor for glaucoma. The model has good predictive performance for glaucoma subtypes, such as AUC=0.74 for primary open-angle glaucoma (POAG); ocular hypertension glaucoma (HTG); AUC=0.73; Normal tension glaucoma (NTG): AUC=0.76 [1].

Andrade De Jesus et al. conducted in-depth research on diagnosis and staging. They scanned angiography (OCTA) based on optical coherence tomography, which included 6 levels of superficial, deep, avascular, panretinal, choroidal capillary layer, choroid, and 7 sectoral areas (Garway-Heath partition). By using machine learning models such as SVM, RF, and xGB, and combining the microvascular density characteristics of multiple layers and regions, the classification is carried out. The study found that OCTA multi-regional characteristics (AUROC ≈ 0.89) were comparable to RNFL thickness (AUROC ≈ 0.85) in glaucoma and healthy control group classifications. In the glaucoma severity classification, OCTA features (xGB model, AUROC = 0.76) were significantly superior to retinal nerve fiber layer (RNFL) (AUROC = 0.67). The most discriminant features are the superficial vascular plexus and the subtemporal region. The disadvantage is that OCTA and RNFL cannot be effectively distinguished between POAG and NTG classifications. Therefore, OCTA multi-layer and multi-regional information has added value in glaucoma severity grading and can be used as a supplement to RNFL [2].

In terms of diagnostic decision-making, Chai et al. developed a Bayesian deep learning model for glaucoma diagnosis, which is also the first Bayesian deep multivariate learning model. By considering uncertainty and integrating information from multiple modalities, such as medical indicators, images, and text, their model achieved better performance in glaucoma detection, providing an information-centric framework for handling multiple medical data sources and addressing decision-making uncertainty in healthcare compared to other methods [3].

Zhang et al. proposed that vision transformers (ViT) and large language models (LLMs) show potential in glaucoma diagnosis, and cross-modal learning can improve model generalization capabilities, such as combining genomic data and electronic health records [4]. In 2023, Li et al. combined clinical data, visual field, and OCT images to predict glaucoma progression within 12 months (AUC = 0.83) [5]; In 2024, Tohye et al. proposed a CA-ViT model that combines GAN enhancement and contour guidance to achieve 93% accuracy in glaucoma classification [6]. Wang et al. used the time series data, the visual field sensitivity sequence in the UWHVF dataset, and the ConvLSTM model to predict the visual field sensitivity of glaucoma patients in the next 0.5~2.0 years. It was found that ConvLSTM performed best in predicting future visual field (MAE = 2.255 dB, $R^2$ = 0.960) and had the best predictive effect on sequence length using the last 3 visual field examinations (1.5~6.0 years). Especially for patients with severe visual field defects, the prediction is more accurate. That is, ConvLSTM can effectively fuse the spatiotemporal features of the field of view sequence, which is better than models that only use temporal or spatial features (such as LSTM and CNN) [7].

Huang et al. contributed to the glaucoma dataset. The GRAPE dataset is a multimodal longitudinal glaucoma dataset containing 1115 follow-up records of 144 patients (263 eyes), covering visual field (VF), color fundus photographs (CFP), OCT measurements and clinical information, and annotating the optic disc/optic cup segmentation and VF progression status for glaucoma management related experiments. The experiment revolves around VF progression prediction and VF estimation. ResNet-

50 was used for VF progression prediction, input baseline CFP, and output progression results based on PLR2, PLR3, and MD slope, with an AUC of 0.80 and an accuracy of 0.91 under the PLR3 standard, with the best performance. VF estimation was also performed using ResNet-50, inputting the original CFP, the ROI of CFP, and the ROI of CFP with segmentation annotation, and the ROI input of CFP was the best, with RMSE 5.475, MAE 4.029, $R^2$ 0.306, and the error distribution was consistent with the pathological characteristics of glaucoma. This dataset fills the gap in the existing glaucoma dataset and promotes the development of glaucoma AI management, but there are limitations such as data selection bias, small sample size, and lack of detailed medication information [8].

## 3. Research on Diabetic Retinopathy Based on Multimodal Information

Retinopathy caused by diabetes is a common eye complication of diabetes that damages retinal blood vessels with long-term hyperglycemia, causing fundus hemorrhage, exudation and other lesions, which can lead to vision loss and even blindness.

Li et al. provided a reliable solution for DR diagnosis in modal information fusion. Experiments were carried out on the diagnosis of diabetic retinopathy with multimodal information fusion. In terms of data, relying on the retrospective and prospective datasets of the EviRed project, covering OCTA and UWF-CFP images of different specifications, and combining supplementary datasets such as GAMMA to verify the robustness of the algorithm. In terms of method, a variety of fusion architectures are proposed: input, single-level, hierarchical fusion and attention mechanism fusion, and the strategy is designed according to different data characteristics, such as the hybrid architecture that combines hierarchy and output fusion for OCTA fusion of different specifications, and the introduction of SE block and manifold hybrid regularization for UWF-CFP and OCTA fusion. The results show that the hierarchical fusion performance is outstanding, with the PDR classification AUC reaching 0.911 in the EviRed retrospective dataset. The hybrid fusion architecture improves the accuracy of multiple classifications, the Kappa value of the six classifications is 0.5593, and the detection AUC of each DR stage is excellent. The detection AUC of the UWF-CFP and OCTA fusion models is higher than that of the single mode in each DR stage, which also verifies the effectiveness of the data processing method [9].

Zhang et al. proposed a multi-model domain adaptive (MMDA) framework for multimodal DR classification to solve the problems of a lack of annotated data and data privacy. The experiment uses DDR, IDRiD, Messidor, and Messidor-2 as the source domain and APTOS 2019 as the target domain, and the preprocessed data are cropped with black borders and image enhancement. ResNet-50 is used as the backbone, combined with the model weight mechanism (MWM), pseudo-label generation and information maximization loss optimization model. At APTOS 2019, MMDA achieved 90.6% accuracy, 98.5% sensitivity, and an AUC of 0.94 using only unlabeled target data and source models, which is lower than some supervised methods, but does not require annotated data. The ablation experiments show that MWM makes the accuracy of the ResNet-50 backbone model 2.6% higher than that of the average weight method, and the performance is optimal when the $\beta$ and $\gamma$ are set to 0.3. t-SNE visualization shows that the model can clearly distinguish between referable and non-referential DR features, providing a solution for low-resource DR screening [10].

In the translation of multimodal medical outcomes into clinical practice, Li et al. developed and verified the multimodal integration system of DeepDR-LLM, which combines a large language model (LLM) based on retinal images and a deep learning module (DeepDR-Transformer) to improve the efficiency and quality of primary diabetes care and diabetic retinopathy (DR) screening. The system has been validated on multiple international multicenter datasets, covering more than 1.24 million standard and portable fundus images. Experiments have shown that DeepDR-Transformer has excellent performance in identifying DR requiring referral (AUC 0.892–0.933), and can significantly improve the diagnostic accuracy (sensitivity increased from 37.2–81.6% to 78.0–98.4%) and efficiency (approximately 23% reduction in reading time) among primary care physicians (PCPs). In

addition, in real-world prospective studies, the management advice provided by PCPs assisted by DeepDR-LLM was better than the unassisted group in terms of quality and empathy, with significant improvements in patient self-management behaviors (e.g., dietary modifications, increased medication adherence), and significant improvements in referral adherence and timeliness. The system shows great potential to promote intelligent and personalized diabetes management in resource-limited areas [11].

## 4. Research on Age-Related Macular Degeneration Based on Multimodal Information

AMD is a common geriatric eye disease with increasing incidence with age, resulting in decreased central vision, visual distortion and even central visual field defects due to degenerative changes in the macular region of the retina.

Wang et al. conducted a multimodal classification of age-related macular degeneration (AMD) and constructed a clinical dataset of 1094 color fundus photographs (CFP) and 1289 optical coherence tomography (OCT) images, covering 1093 eyes, divided into four categories: normal, dry AMD, polypoid choroidal vascular disease (PCV), and wet AMD [12]. A dual-stream CNN architecture is proposed, with ResNet-18 as the backbone, using SI-Fusion to combine CFP and OCT modal information, and extended class activation mapping (CAM) to achieve multi-modal contribution visualization. In order to solve the shortcomings of multimodal data, two data enhancement methods are designed: CAM conditional image synthesis based on GAN, which uses CAM as pix2pixHD input to generate high-resolution CFP/OCT images; Loose Pairing pairs images by category rather than eye identity. Experiments show that the F1 value of dual-flow CNN (MM-CNN-da) reaches 0.914 and the accuracy is 0.863 in the four classification tasks, which is significantly better than the unimodal baseline (OCT-CNN F1 0.886, accuracy 0.818) and the traditional multimodal method (Yoo et al. method F1= 0.792, accuracy = 0.690) [13], and effectively reduces the misclassification rate of PCV and wet AMD, verifying the effectiveness of multimodal fusion and data enhancement methods [12].

Suresh's study focused on the management of AMD in the elderly population, combining clinical cases with community screening. In clinical cases, a 65-year-old male patient with dry AMD was treated with intravitreal (IVI) broluluzumab with an interval of 20 weeks and good visual and anatomical results. A 79-year-old female patient with wet AMD was treated with (Lucentis) anti-VEGF and her vision stabilized. Community screening was conducted in 95 people aged 60 years, and 91 images were gradable, of which 26 were diabetic and 65 had diabetes. A total of 20 cases of AMD were detected, 14 of which were early-stage and 6 patients required anti-VEGF injection (2 with broluzumab, 2 with, and 2 with bevacizumab). Studies have found that diabetic retinopathy (DR) has a protective effect on AMD, which is often accompanied by diabetes, elevated systolic blood pressure, high BMI, and high triglycerides. In addition, retrospective studies have shown a prevalence of AMD in India of 1.2%-4.7%, cataract surgery is associated with an increased risk of AMD, and DR and glaucoma reduce the risk of AMD. Anti-VEGF drugs (e.g., broluzumab) for the treatment of neovascular AMD (nAMD) are treated with a treatment-on-demand (PRN) regimen in most patients, with an average injection-free interval of 19.43 weeks, and a good safety profile, providing a practical basis for the clinical management of AMD [14].

## 5. Challenges and Prospects

### 5.1. Challenges

In general, the integration and complementary mining of existing multimodal technologies still need to be explored in depth. Although OCT, fundus photography, and visual field examination can be integrated, the spatiotemporal resolution registration errors of each modality are large, such as OCT static high-resolution structural data and visual field dynamic low-resolution functional data are difficult to accurately correlate with optic nerve damage and visual field defects. CFP and OCT were

fused, and clinical data such as blood glucose were not included, and OCTA and CFP characteristics were not fully integrated. The dual-flow CNN model had a 15% false positive rate for PCV vs. wet AMD. At the same time, there is also the problem of personality. For example, the OCTA and RNFL thickness fusion models in glaucoma diagnosis and treatment have limited effectiveness in distinguishing between NTG and POAG, and AUROC has not exceeded 0.7. The data quality also needs to be improved, and the follow-up data equipment is very different, the annotation standards are not uniform, and most of them are single-center data, lacking diversity. It is difficult to implement clinical practice, and the "black box" characteristics of the model make it difficult for doctors to understand the diagnostic logic, and the trust between man and machine needs to be improved. In addition, DR lesions are difficult to label, and the manifestations of different modalities vary greatly, and the judgment of "intermediate lesions" is very different, and PDR neovascularization is easily confused with artifacts. Models are mostly used for diagnosis, lack the ability to evaluate treatment response, and it is difficult to quantify the association between post-treatment images and vision. The diagnosis and treatment mode of dry AMD is short, and the multimodal model lacks FAF and OCT joint analysis, making it difficult to predict the progression of GA. Insufficient treatment monitoring makes it difficult to quantify the association between post-treatment imaging and visual acuity, which cannot explain the contradiction in efficacy.

### 5.2. Prospects

In terms of data scarcity, a deep learning multimodal spatial registration algorithm can be developed, multi-task learning can be introduced, and dry AMD samples can be supplemented with GANs. Integrate FAF, OCT and genetic data to build a risk prediction model for dry AMD, explore OCTA applications to build a longitudinal database, develop treatment response and prognosis prediction models, and combine wearable devices to achieve dynamic correlation.

In the future, a dynamic fusion model of "structure-function-vessels" can be developed on the model, combined with multimodal data and time series modeling, a multi-center database can be designed with reference to longitudinal data, and explainable AI can be introduced. Promote the standardization of data collection and annotation, use federated learning to integrate multi-center data, and build personalized models based on genetic data. Develop portable multi-modal devices, with lightweight models, and combine wearable devices to build an all-round monitoring system. Construct an "imaging-clinical-molecular" multimodal system, integrate multiple types of data, and use the transformer architecture to improve the diagnosis rate of complex lesions. Generative AI is used to synthesize samples, develop multimodal lesion tracking algorithms, and build treatment response prediction models. Develop low-cost portable devices, with lightweight models, combined with 5G and remote expert systems, to solve grassroots diagnosis and treatment problems.

## 6. Conclusion

This article reviews the application progress and value of multimodal medical technology, focusing on three types of blinding eye diseases: glaucoma, diabetic retinopathy (DR) and age-related macular degeneration. The results show that multimodal technology significantly compensates for the lack of single modality by integrating OCT, OCTA, CFP and other imaging data, as well as clinical and genetic information: in the diagnosis and treatment of glaucoma, the combination of OCTA multi-regional features and RNFL improves the grading accuracy, the ConvLSTM model accurately predicts the progress of visual field, and the GRAPE dataset fills the gap in longitudinal research. In DR diagnosis and treatment, the hierarchical fusion architecture, MMDA framework, and DeepDR-LLM system optimize diagnostic accuracy, solve data scarcity problems, and improve grassroots diagnosis and treatment efficiency, respectively. In the diagnosis and treatment of AMD, the dual-flow CNN model combined with data enhancement to improve the classification effect, and the efficacy of anti-VEGF drugs was clear and DR had a protective effect on AMD.

However, this field still faces challenges, with large registration errors in multimodal data, prominent disease-specific challenges, uneven data quality, and model "black box" characteristics hindering clinical implementation. In the future, it is necessary to make breakthroughs in technology, data, and clinical transformation, develop dynamic fusion models and explainable AI, build multi-center standardized databases, develop portable devices and lightweight models, promote the widespread application of multimodal technology in personalized and intelligent diagnosis and treatment of ophthalmic diseases, and ultimately improve the level of global ophthalmic medical services.

## References


[1] Omodaka K, An G, Tsuda S, et al. Classification of optic disc shape in glaucoma using machine learning based on quantified ocular parameters. PloS One, 2017, 12 (12): e0190012.

[2] Andrade De Jesus D, Sánchez Brea L, Barbosa Breda J, Fokkinga E, Ederveen V, Borren N, Bekkers A, Pircher M, Stalmans I, Klein S, van Walsum T. OCTA multi-layer and multi-sector peripapillary microvasculature modeling for glaucoma diagnosis and staging. Translational Vision Science & Technology, 2020, 9 (2): 58. Available at: https://doi.org/10.1167/tvst.9.2.58

[3] Yidong C, Yiyang B, Hongyan L, et al. Glaucoma diagnosis in the Chinese context: An uncertainty information-centric Bayesian deep learning model. Information Processing and Management, 2021, 58 (2): 102454.

[4] Zhang J, Tian B, Tian M, et al. A scoping review of advancements in machine learning for glaucoma: Current trends and future direction. Frontiers in Medicine, 2025, 12: 1573329.

[5] Li Y, Han YJ, Li ZH, Zhong Y, Guo ZF. A transfer learning-based multimodal neural network combining metadata and multiple medical images for glaucoma type diagnosis. Scientific Reports, 2023, 13: 13.

[6] Tohye TG, Qin Z, Al-Antari MA, Ukwuoma CC, Lonseko ZM, Gu YH. Ca-ViT: Contour-guided and augmented vision transformers to enhance glaucoma classification using fundus images. Bioengineering, 2024, 11: 23.

[7] Wo W, Zheng Xiujuan, Lü Zhiqing, et al. Visual field prediction research based on spatiotemporal feature learning. Journal of Biomedical Engineering, 2024, 41 (5): 1003.

[8] Huang X, Kong X, Shen Z, et al. GRAPE: A multi-modal dataset of longitudinal follow-up visual field and fundus images for glaucoma management. Scientific Data, 2023, 10 (1): 520.

[9] Li Y, Hajj HA, Conze PH, et al. Multimodal information fusion for the diagnosis of diabetic retinopathy. arXiv preprint arXiv:2304.00003, 2023.

[10] Zhang G, Sun B, Zhang Z, et al. Multi-model domain adaptation for diabetic retinopathy classification. Frontiers in Physiology, 2022, 13: 918929.

[11] Li J, Guan Z, Wang J, et al. Integrated image-based deep learning and language models for primary diabetes care. Nature Medicine, 2024, 30 (10): 2886-2896.

[12] Wang W, Li X, Xu Z, et al. Learning two-stream CNN for multi-modal age-related macular degeneration categorization. IEEE Journal of Biomedical and Health Informatics, 2022, 26 (8): 4111-4122.

[13] Yoo TK, Choi JY, Seo JG, Ramasubramanian B, Selvaperumal S, Kim DW. The possibility of the combination of OCT and fundus images for improving the diagnostic accuracy of deep learning for age-related macular degeneration: A preliminary experiment. Medical & Biological Engineering & Computing, 2019, 57 (3): 677-687.

[14] Suresh K. Managing age-related macular degeneration (AMD) among elderly: Early diagnosis, lifestyle changes & anti-VEGF drugs ensure better outcomes. JOJ Ophthalmology, 2025, 12 (5): 555850.