

基于眼底照相及 OCT 的多模态 AI 在眼科疾病中的应用与临床价值

柴 爱^{1,2}, 余 萍²

引用:柴爱,余萍. 基于眼底照相及 OCT 的多模态 AI 在眼科疾病中的应用与临床价值. 国际眼科杂志, 2026, 26 (10): 1772-1777.

作者单位:¹(810000)中国青海省西宁市,青海大学;²(810000)中国青海省西宁市,青海大学附属医院眼科
作者简介:柴爱,女,在读硕士研究生,住院医师,研究方向:眼底病。
通讯作者:余萍,女,硕士研究生,主任医师,硕士研究生导师,研究方向:屈光不正、眼表疾病、青光眼、白内障、眼底病。1156177284@qq.com
收稿日期:2026-03-28 **修回日期:**2026-08-12

摘要
近年来,人工智能(AI)在眼科的应用范围持续扩展,AI 辅助分析彩色眼底照相(CFP)和光学相干断层扫描(OCT),可以获取更多的视网膜血管量化数据,为评估全身血管状态提供帮助。单一模态 AI 系统(仅基于 CFP 图像或仅基于 OCT 图像)虽取得一定成效,但无法模拟临床医生整合多源信息进行综合判断的思维过程。多模态 AI 融合 CFP 的结构信息和 OCT 的层间信息,能够挖掘不同模态的互补特征,提高诊断准确性和预测能力。文章系统综述基于 CFP 和 OCT 的多模态 AI 在致盲性眼病中的应用进展,阐述其技术基础,以及在糖尿病视网膜病变、年龄相关性黄斑变性、青光眼、病理性近视等致盲性眼病中的应用现状及临床价值,并讨论当前多模态 AI 面临的挑战及未来发展方向,为眼科医生提供参考依据,推动眼科诊疗技术的发展。
关键词:多模态人工智能(AI);彩色眼底照相(CFP);光学相干断层扫描(OCT);深度学习;眼科疾病
DOI:10.3980/j.issn.1672-5123.2026.10.15

Application and clinical value of multimodal AI based on fundus photography and OCT in ophthalmic diseases

Chai Ai^{1,2}, Yu Ping²

¹Qinghai University, Xining 810000, Qinghai Province, China;
²Department of Ophthalmology, Qinghai University Affiliated Hospital, Xining 810000, Qinghai Province, China
Correspondence to: Yu Ping. Department of Ophthalmology, Qinghai University Affiliated Hospital, Xining 810000, Qinghai Province, China. 1156177284@qq.com
Received:2026-03-28 Accepted:2026-08-12

Abstract

• In recent years, the application of artificial intelligence (AI) in ophthalmology has been continuously expanding. AI-assisted analysis of color fundus photography (CFP) and optical coherence tomography (OCT) enables the acquisition of more quantitative data on retinal vasculature, providing assistance in evaluating the systemic vascular status. Although single - modal AI systems (based solely on fundus analysis or OCT images) have achieved certain success, they are unable to simulate the comprehensive judgment process of clinicians who integrate multi - source information. Multimodal AI, which combines structural information from CFP with interlayer information from OCT, can uncover complementary features across different modalities, thereby enhancing diagnostic accuracy and predictive capabilities. This paper provides a systematic review of the application progress of multimodal AI based on CFP and OCT in blinding eye diseases. It outlines the technical foundations, current applications, and clinical value of multimodal AI in conditions such as diabetic retinopathy, age - related macular degeneration, glaucoma, and pathological myopia. Furthermore, the challenges and future directions of multimodal AI are discussed, offering references for ophthalmologists and promoting the development of ophthalmic diagnosis and therapeutic technologies.
• **KEYWORDS:** multimodal artificial intelligence (AI); color fundus photography (CFP); optical coherence tomography (OCT); deep learning; ophthalmic diseases

Citation: Chai A, Yu P. Application and clinical value of multimodal AI based on fundus photography and OCT in ophthalmic diseases. Guoji Yanke Zazhi (Int Eye Sci), 2026, 26 (10): 1772-1777.

0 引言

眼科是高度依赖影像学检查的学科,早期诊断和精准治疗对于预防视力损害至关重要。随着人口老龄化加快,糖尿病视网膜病变(diabetic retinopathy, DR)、年龄相关性黄斑变性(age-related macular degeneration, ARMD)、青光眼等致盲性眼病的患病率持续上升。彩色眼底照相(color fundus photography, CFP)和光学相干断层扫描(optical coherence tomography, OCT)是眼科最常用的两种影像学检查手段。CFP 提供视网膜表面的二维彩色图像,直观显示眼底宏观病变;OCT 通过高分辨率获取视网膜的横断面图像,定量分析分层结构。两种技术信息互补,联合使用评估眼部病变来提高诊断准确性。然而,传统诊断高度

依赖医生的经验,临床工作负荷增加与专家短缺,导致报告判读出现延迟^[1]。基于单一影像模式的人工智能(artificial intelligence, AI)系统已取得显著进展,而多模态AI可以通过整合不同来源的医学信息,构建更全面的疾病特征。在眼科领域,多项研究开发并验证了AI算法在视网膜成像上的应用,包括DR、ARMD、青光眼及其他视网膜疾病的诊断和管理^[2-5]。近年来,基于OCT和CFP多模态数据训练的AI模型在病灶检测和定量分析方面表现优异^[6]。本文系统综述基于CFP和OCT的多模态AI在眼科疾病中的应用现状与临床价值,为该领域研究提供参考。

1 多模态 AI 的技术基础

1.1 多模态融合策略 多模态AI的核心在于有效整合来自不同模式的异构数据,实现对共同现象的不同视角分析,达到“1+1>2”的效果。目前主要采用三种融合策略:(1)早期融合(early fusion, EF):是一种基础方法,指在完成各模式数据的初步特征提取与预处理之后,在数据层面直接进行拼接、连接或组合,形成统一的高维特征向量,并输入至建模,进入预测流程。该方法的核心优势在于能充分保留各模式的原始语义信息,尤其适用于不同模式在特征维度、数据尺度相对一致的场景中^[7]。(2)晚期融合(late fusion):又称决策层融合,该策略不直接对原始数据或特征进行合并,而是为每个模式分别构建独立的子模型,在各特征空间内完成训练与预测,最后通过加权平均、投票机制、集成学习算法等方式,将多个子模型的输出结果整合为最终决策。该方法核心优势在于具有较高灵活性和鲁棒性,可处理模式缺失,便于集成现有单模式模型。(3)混合融合(hybrid fusion):结合了早期与晚期融合的特点,在不同阶段进行多层次特征交互与融合,以兼顾特征互补性与模型可解释性,可挖掘更深层的扩模式关联。多模态AI融合策略对比见表1。

1.2 关键算法架构 卷积神经网络(CNN):利用卷积算子直接从图像中提取层次化特征,在图像分类任务中非常有效。图像经过卷积层、池化层和激活层处理以提取特征,最终经过全连通层对图像进行分类。视觉语言模型(vision languagemodels, VLMs):这是一种新型AI范式,能够在统一框架内中同时分析图像和文本数据。通过自然语言描述目标属性,在共享特征空间中建立视觉特征与语义描述的映射关系,缓解类内差异大、类间差异小的挑战。基础模型(foundation models):是将计算机视觉与自然语言处理相结合的AI模型,旨在联合视觉和文本信息进行理解,为下游具体任务提供支持。通过通用架构设计、预训练与微调的领域,形成专用视觉语言模型。

2 多模态 AI 在眼科疾病中的应用

2.1 DR 糖尿病(diabetes mellitus, DM)以高血糖为特征,长期高血糖状态会对多个器官造成严重损害,包括心脏、肾脏和视网膜等^[8-9]。DR是DM最常见、最严重的并发症之一^[10]。传统的DR筛查方法主要为CFP,对黄斑水肿(diabetic macular edema, DME)的判断能力有限,而OCT可清晰显示DME所在的视网膜形态改变,是DME诊断的金标准^[11-12]。用于视网膜图像分析的AI系统的发展,为开发非侵入性、可扩展且经济高效的糖尿病系统性并发症检测创造机会。

多模态AI通过联合CFP与OCT,实现对DR病变程度和DME的综合评估。既往研究仅使用单一成像模式(CFP或OCT),而结合多模态可能提升预测能力^[13]。Gulshan等^[14]以Inception-v3为架构,使用10万余张CFP图像进行训练,在不同数据集上的曲线下面积(AUC)分别为0.963和0.892,识别DR和DME的敏感度和特异度在88.9%-97.5%,模型的性能丝毫不逊色于眼科专家^[15]。Gu等^[16]利用了1445对CFP-OCT配对图像,开发了Fusion-MIL、CFP-MIL和OCT-MIL模型,Fusion-MIL在多个测试数据集中平均AUC达到0.962-0.985,明显优于CFP-MIL及OCT-MIL,可见双模式融合模型可提升诊断水平,实现更高精度。也有研究表明,多模态(CFP和OCT)模型并未比单独的OCT模型带来额外优势^[17]。在其主要测试集中,CFP模型、OCT模型和多模态模型分别检测出伴视觉障碍的DME,分别是0.756-0.949、0.756-0.949、0.907-0.939,多模态AI模型中表现良好。自动化报告生成系统整合CFP和OCT发现,多模态AI提高了DME的筛查和检测灵敏度、特异性,提升了临床效率和患者预后^[18-19]。

2.2 ARMD ARMD是一种进行性视网膜疾病,主要影响负责中心视力的黄斑区。ARMD的早期和中期阶段可能无明显症状,但晚期ARMD可导致严重的视力障碍,影响视觉敏锐度^[20]。ARMD分为干性和湿性两类,湿性ARMD需及时抗VEGF治疗,而治疗决策依赖于对CFP(玻璃膜疣、地图样萎缩)和OCT(视网膜下积液、视网膜内液、脉络膜新生血管等)综合信息的判断^[21]。在眼底病领域,AI可通过CFP和OCT图像对ARMD做出诊断,准确率可达眼科专家水平。

湿性ARMD的主要临床体征为DME与脉络膜新生血管(choroidal neovascularization, CNV)。通过深度学习(deep learning, DL),AI可以自动检测水肿体积、色素上皮层下积液、神经上皮层脱离来诊断DME及CNV^[22]。地图样萎缩病变(geographic atrophy, GA)是干性ARMD变性的

表 1 多模态 AI 融合策略对比

融合类型	核心思路	优势	代表方法	适用场景
早期融合	在数据/特征输入层直接拼接形成统一的高维特征向量	充分保留各模式的原始语义信息	特征拼接、卷积融合、共享底层编码器	不同模式在特征维度、数据尺度相对一致的场景中
晚期融合	各模式独立完成训练、预测,在决策层进行融合	较高灵活性、鲁棒性,可处理模式缺失,便于集成现有单模式模型	加权、投票、平均、逻辑回归融合	不同模式数据来源、多个单模式模型的集成
混合融合	多阶段组合,在不同层次特征交互与融合	兼顾特征互补性与模型可解释性,挖掘更深层的扩模式关联	多级注意力、分层融合、残差门控	追求最佳性能,大规模高质量的配对数据的复杂场景

晚期阶段,目前已有多种模型通过多种成像特征可预测GA的进展速度。多数研究基于CFP或OCT应用DL对ARMD进行分析、分类。Khalid等^[23]通过建立OCT和CFP对应关系识别ARMD的模型。该模型在OCT图像分析技术上可以达到96.4%、97.1%、96.19%的准确性、灵敏度和特异性,在同一数据集上的CFP分析分别达到了86%、76.7%和90%,当分析CFP和OCT的融合系统时,达到了98%、100%和97.14%,可以看出融合模型明显高于单模型。Yoo等^[24]通过结合CFP和OCT的DL模型对ARMD进行分类,仅使用OCT的DL诊断平均准确率达82.6%,仅使用CFP表现出83.5%的准确率,而将CFP和OCT结合可提高平均准确率达90.5%。上述研究表明,CFP和OCT成像可在视网膜上提供互补的信息,二者结合的DL模型具有更好的效能。

2.3 青光眼 青光眼主要与病理性眼压升高相关,损害视神经及视觉通路,最终导致视觉功能丧失。其典型表现是视盘萎缩和视野缺损及缩小。该疾病早期通常无明显症状,周边视力也无明显下降,很多患者在疾病的晚期才来就诊,此时视功能损伤已不可逆转。传统OCT报告需要将视网膜神经纤维层(retinal nerve fibre layer,RNFL)厚度与OCT中内置的正常人数据库进行比较,但临床上无用于诊断或排除青光眼的RNFL厚度特定阈值。此外,RNFL厚度测量受图像质量差、患者配合不佳、近视等多种因素影响^[25-26]。

此前青光眼检测模型主要集中于单一模态^[27-31],现通过整合多模态数据,提升模型性能并更好地契合临床实践。多模态AI可以通过CFP和OCT同时获得视盘参数(杯盘比、盘沿宽度等参数)、RNFL厚度、神经节细胞复合体厚度,构建联合模型。基于DL的医学图像分析系统在疾病检测准确性与临床医生相当甚至更优,能够更准确、更快速地检测青光眼及其变化^[32-34]。Hwang等^[35]基于多模态神经网络青光眼检测(MGD)模型,对多模态图像进行训练和评估。结果显示多模态集成可显著提升性能指标(AUC:0.86,特异性:0.85),可以看出综合性方法兼顾结构和功能数据,显著提升AI的功能性和准确性。Chuter等^[36]通过CFP和OCT图像开发了多模态RETFound模型,与仅使用单模态RETFound模型进行了比较,多模态模型实现了更高的AUC、准确率、回忆率(0.94、0.92、0.69),显著优于CFP单模模型,但未超过OCT单模模型(AUC:0.93),这可能与图像数量存在差异有关。DL的进步明显提升了通过OCT成像的青光眼进展评估,这些方法在提升诊断准确性、监测疾病进展以及为青光眼病理生理展现巨大潜力。

2.4 高度近视与病理性近视 高度近视指屈光度 ≥ -6.00 D,以眼轴增长为特点,会引起视乳头变化、视网膜劈裂、黄斑裂孔及后巩膜葡萄肿等并发症,从而导致不可逆的视网膜细胞损伤,造成视功能降低^[37]。病理性近视(pathologic myopia, PM)是高度近视的进展阶段,以后巩膜葡萄肿和继发性眼底退行性病变为特征,伴有不可逆的视力损害。

OCT成像被证明是比CFP更准确的诊断工具,能够精确评估疾病进展。然而,由于时间和成本高,OCT并非首选检查,这凸显了利用CFP、OCT和超广角眼底成像(Ultra-widefield,UWF)多模态成像在临床环境中协同运作的需求。高度近视眼底病变的分类主要采用ATN分级

系统,是眼底图像进行标签分类的重要依据。潘新蕾^[38]通过采用ResNet50的卷积模块构建多模态学习模型对PM进行研究,在识别及判断需要治疗的患者方面,多模态模型表现较为优异,其准确度在0.842-0.975,AUC均高于0.925,而单模态模型的准确度在0.713-0.933,AUC在0.658-0.824,均低于多模态模型。Xu等^[39]在PM中对单模态CFP、OCT和双模态AI进行了ATN评分的表现比较。双峰模型—双重DL在分类中表现出优越的准确性。多模态AI在治疗重度PM的AUC为0.9635,而OCT模型为0.9359,CFP模型为0.9268。可见单纯依靠CFP或OCT判断PM患者的诊疗方式并不可靠,需结合CFP和OCT图像的双模态AI来识别需行处理的患者,并合理判断其治疗方法。

2.5 其他类型视网膜疾病 视网膜静脉阻塞(RVO)是第二常见的视网膜血管疾病。视网膜分支阻塞(BRVO)以中央视网膜静脉主要分支之一闭塞为特征,视力障碍严重程度主要取决于闭塞的位置、范围和持续时间。Choi等^[40]研究通过多模态方法,构建了仅使用眼底图像的单峰模型、视网膜视盘和血管增强多模态模型以预测未来的BRVO,结果显示多模态模型优于单模模型,AUC为0.76,准确率为68.5%。这些发现展示了视网膜视盘和血管增强多模态方法在BRVO预测中的潜力,BRVO预测性能显著提升。Barbosa等^[41]使用VGG-19网络开发了三种模型:单模态(仅为OCT B扫描)、多模态(OCT的B扫描和糖尿病信息)、多图像(OCT的B扫描、红外图像和糖尿病信息)。结果显示多图像模型取得了更好的结果,准确率为95.20%,精度为95.43%,召回率为95.20%,AUC为99.59%。多模态AI在BRVO中展现出巨大潜力。

中央浆液性绒毛膜视网膜病变(CSCR)是一种特发性黄斑疾病,其特征是视网膜色素上皮(retinal pigment epithelium,RPE)功能障碍和脉络膜高透性,导致神经上皮层脱离(neurosensory retinal detachment,NRD)和/或色素上皮脱离(pigment epithelium detachment,PED)。OCT与CFP可视化相结合,使CSCR能够可靠诊断,并有效区分其他血管病变(如DR、湿性ARMD)^[42]。Xu等^[43]引入多模态图像融合方法,开发了迭代学习模型(ILM-GRESS和ILM-SALML)框架,实现了病灶直径的准确测量(修正率为83.0%)。整合CFP和OCT数据的病灶边界定位模块(LBLM)将校正率提升至92.11%。Yoo等^[44]提出了结合迁移学习与多模态数据(CFP和OCT/临床参数)预测1a光动力疗法结局(敏感与耐药)。该综合模型的准确率达到88%,显著优于仅基于成像的模型。AI在CSCR中的应用已从疾病分类发展到动态预后预测,利用多模态数据融合(如CEP、OCT、OCTA)来提升诊断准确性。多模态AI在眼病中的应用见表2。

3 从眼科到全身的拓展应用

视网膜作为中枢神经系统的延伸,是全身疾病诊断的重要窗口。当视网膜OCT、CFP与DL算法相结合,可以用于监测许多神经退行性疾病的进展,如多发性硬化症、阿尔茨海默病(Alzheimer's disease, AD)^[45]和帕金森病(Parkinson's disease,PD)^[46]、中风等疾病。视网膜可以直接、无创地观察和分析血管的部位,通过评估视网膜血管损伤的严重程度,间接反映外周、大脑和冠状血管的血管疾病。Tan等^[47]汇总多项研究数据,开展了关于OCT在全身性高血压应用的荟萃分析。Jiang等^[48]通过荟萃分析

表 2 多模态 AI 在眼病中的应用

疾病	单模态(仅 CFP/仅 OCT)	多模态(CFP+OCT)	核心性能指标(融合模型)	参考文献
DR	CFP 对 DME 的判断有限,OCT 无法提供视网膜血管病变	综合评估 DR 病变程度和 DME 状态	AUC 达 0.907–0.985,敏感度、特异度达 88.9%–97.5%	Gu 等 ^[16] 和 Tan 等 ^[17]
ARMD	无法对玻璃膜疣、GA、CNV、视网膜下积液等综合判断	实现液体渗出的自动检测和量化系统	准确度、灵敏度、特异性达 98%、100%、97.14%	Khalid 等 ^[23]
青光眼	RNFL 厚度临床上无特定阈值。测量受多因素影响	同时获得视盘参数、RNFL 厚度、神经节细胞复合体厚度	AUC 达 0.94,优于 CFP 单模模型(AUC:0.86)	Chuter 等 ^[36]
PM	无法精准采用 ATN 分级系统,对各种并发症研究有限	多模态通过 ATN 分级,对视力预后进行精确评估	准确度 0.842–0.975,AUC 高于 0.925	潘新蕾 ^[38]
RVO	仅依赖单一设备图像限制了普适性	结合 CFP 和视网膜视盘和血管综合分析	AUC 达 0.76,准确率达 68.5%	Choi 等 ^[40]
CSCR	单模态无法有效区分其他血管病变	多模态数据可提高 CSRS 诊断性能	准确率达到 88%,显著优于单模态模型	Yoo 等 ^[44]

研究评估了 SD – OCT 对多足型脉络膜血管病变 (polypoidal choroidal vasculopathy, PCV) 的诊断敏感性和特异性。Pellegrini 等^[49]阐述了 OCTA 在神经退行性疾病中的应用。Yoo 等^[24]讨论了 AI 在全身健康和长寿医学中的应用,为眼科成像领域的做出重要贡献。

4 多模态 AI 的临床价值

4.1 提高诊断准确性 多模态 AI 通过整合 CFP 与 OCT 互补信息,提高疾病检出率和分类准确性。多项研究表明,与单模态相比,多模态系统 AUC 提升 4%–5%,准确度提升 2%–7%^[19]。这种融合模式有效弥补单一模态影像的固有短板,显著降低了诊断不确定性及假阴性率。

4.2 提升临床效率 自然语言处理(NLP)和多模态 AI 的结合推动自动化报告生成的临床应用。多模态 AI 系统通过整合多样化数据源,加快临床工作流程,减少医师阅片时间,提高工作效率,显著降低漏诊率和转诊延误。

4.3 标准化随访方案 在分级诊疗场景中,多模态 AI 展现出从筛查到转诊,AI 辅助诊断可减少不必要的专科转诊,优化医疗资源配置。AI 生成的眼科报告有望从回顾性文档演变为动态、预测性和患者交互式工具,有助于降低医疗成本,可减少晚期并发症的治疗费用。

5 现有研究争议与辩证分析

多数研究针对 CFP 联合 OCT 多模态 AI 展现出优于单模态的诊断性能,但部分研究结论存在明显分歧。在 DR、ARMD、PM、RVO 中多数研究证实融合模型性能提升:Gu 等^[16] DR 双模态 Fusion–MIL 模型 AUC 为 0.962–0.985,显著高于单独 CFP/OCT 模型;Khalid 等^[23] ARMD 融合模型准确度 98%、灵敏度 100%,远超单模态。而 Tan 等^[17]研究伴视力损伤的 DME 时发现,多模态模型 AUC(0.907–0.939)并未显著超越单模型 OCT,甚至在某些子集中略逊于 OCT;同样,Chuter 等^[36]青光眼 RETFound 基础模型试验中,多模态融合模型 AUC 为 0.94,仅优于 CFP 单模,但未超过 OCT 单模态模型,融合未实现性能跃升。

深入分析分歧根源可见,DME、青光眼的诊断依据黄斑分层厚度、RNFL 等 OCT 量化指标,CFP 仅提供辅助宏观血管信息,补充价值有限,甚至因配准误差或图像质量不匹配而引入噪声,从而抵消多特征融合带来的性能提升。而合并大范围血管渗出、玻璃膜疣、后巩膜葡萄肿的眼底病,CFP 和 OCT 包含独立诊断价值,融合增益显著。

其次,Chuter 等^[36]研究中 CFP–OCT 配对图像数量不均衡,部分低质量 CFP 图像引入噪声,抵消了融合优势;而 Gu 等^[16]、Khalid 等^[23]采用标准化设备采集、高质量配对图像,无模态数据噪声干扰,融合优势充分释放。最后,采用晚期决策融合的青光眼模型易出现“单 OCT 模型权重过高”,CFP 特征被淹没;早期特征拼接融合的 DR、ARMD 模型可平等提取两类图像特征,互补作用充分发挥。融合策略不匹配疾病影像特征,简单融合可能导致梯度冲突或特征淹没,会直接造成“多模态无提升”的试验假象,削弱模型性能。

综上,多模态 AI 的性能优势是条件性和策略敏感性的,并非普适性优势。现有研究结论冲突由疾病病理特征、数据集构建、融合算法等多因素导致,当前研究仍缺乏分层、多中心及标准化对照试验。未来需开发自适应权重融合架构,整合多中心异质影像数据,分层区分不同眼病、不同基层诊疗条件开展验证,从而实现多模态 AI 真实临床价值。

6 当前临床诊疗的影响及未来发展趋势

6.1 多模态 AI 对当前眼科临床诊疗的影响 传统眼科治疗决策依赖指南推荐的标准方案(如 DR 的定期筛查、ARMD 的抗 VEGF 按需治疗),但启动治疗的个体化阈值高度依赖医师主观阅片。多模态 CFP+OCT 的融合模型整合宏观眼底血管、视网膜分层量化双重特征,实现客观、量化、标准化疾病分层。

在 DR、DME 中,多模态 AI 可同步完成 DR 严重程度分级及黄斑水肿体积自动测算,精准区分无需干预轻度 DR、需激光光凝的重度非增殖期 DR、需抗 VEGF 注射的合并中心累及 DME 患者;在湿性 ARMD 中,AI 自动量化视网膜下积液,脉络膜新生血管范围,替代传统人工粗略评估,提前检测到活动性病变,从而指导抗 VEGF 治疗的提前启动。在青光眼中,多模态 AI 融合 CFP 杯盘比、盘沿形态与 OCT 神经节细胞复合体、RNFL 厚度,区分进展型青光眼与稳定型青光眼,指导降眼压药物、激光小梁成形术、微创青光眼手术分层干预。在 PM 中,多模态 AI 自动完成 ATN 量化评分,精准识别需黄斑裂孔修复、后巩膜加固术的高危患者,降低单纯依靠 CFP 低估视网膜劈裂、CNV 的漏治风险。同时多模态 AI 可将不同时间点的 CFP 与 OCT 图像进行空间对齐,量化病灶体积、液体量、RNFL 厚度的动态变化速率。这种“轨迹预测”功能可早期识别

治疗效果不佳者,从而及时调整方案,避免无效治疗的持续。

6.2 多模态 AI 未来临床发展趋势 结合多模态 AI 现阶段临床应用价值、局限性及现有研究分歧,未来多模态 AI 应聚焦于以下方向:(1)突破单一疾病专用模型局限,构建可兼容 CFP、OCT、光学相干断层扫描血管成像(optical coherence tomography angiography,OCTA)、临床病历文本的多模态基础模型(EyeFM),同时引入注意力动态权重混合融合架构,自动识别核心病灶,针对以 OCT 层结构损伤为主的青光眼,以血管病变为主的 DR、RVO,平衡双模态信息输入,从算法层面解决“多模态是否优于单模态”的试验结论冲突。(2)由静态诊断转向动态监测,开发长期纵向影像配准、病灶差值量化功能,自动存储、比对患者历年 CFP 与 OCT 数据,精准量化积液体积、RNFL 厚度、萎缩病灶面积的动态变化,实现个体化预后预测与治疗反应评估。(3)开发可解释 AI,克服“AI 黑箱”挑战,让 AI 推理过程透明化,通过病灶热力图、分层分割可视化解决深度学习“黑箱”缺陷,同时制定统一的伦理框架,推动多模态 AI 辅助诊疗方案纳入眼底病临床指南,获得临床医生的信任及促进监管审批。(4)依托联邦学习整合多中心、多设备异质数据,消除单中心数据偏倚,建立无需共享患者原始影像即可联合训练多模态模型,解决数据孤岛和隐私问题,统一全国 AI 诊断标准。

7 结论

基于 CFP 和 OCT 的多模态 AI 是当前眼科诊断技术的重要发展方向。通过融合宏观形态与微观量化的影像信息,多模态 AI 系统在 DR、ARMD、青光眼、高度近视及其他致盲性眼病的筛查、诊断和监测中展现出显著的临床价值,在 RVO 等复杂场景中展现出超越人类的预后预测能力。从眼科向全身疾病的拓展,多模态 AI 有望成为链接眼部健康与系统性疾病管理的桥梁。随着基础模型、联邦学习、可解释 AI 及多中心研究的进步,实现从“辅助诊断”到“全流程管理”的升级。在规范数据,完善伦理的基础上,多模态 AI 将在精准眼科医疗中发挥越来越重要的作用,为全球视力健康事业做出更大贡献。

利益冲突声明:本文不存在利益冲突。

作者贡献声明:柴爱论文选题与修改,初稿撰写,文献检索,数据分析;余萍选题指导,论文修改及审阅。所有作者阅读并同意最终的文本。

参考文献

[1] Wang SY. The rural shortage of ophthalmic subspecialists. *JAMA Ophthalmol*, 2025,143(2):125.
[2] Abramoff MD, Whitestone N, Patnaik JL, et al. Autonomous artificial intelligence increases real-world specialist clinic productivity in a cluster-randomized trial. *NPJ Digit Med*, 2023,6(1):184.
[3] Grzybowski A, Brona P, Lim G, et al. Artificial intelligence for diabetic retinopathy screening: a review. *Eye*, 2020,34(3):451-460.
[4] Hood DC, La Bruna S, Tsamis E, et al. Detecting glaucoma with only OCT: Implications for the clinic, research, screening, and AI development. *Prog Retin Eye Res*, 2022,90:101052.
[5] Dong L, He WJ, Zhang RH, et al. Artificial intelligence for screening of multiple retinal and optic nerve diseases. *JAMA Netw Open*, 2022,5(5):e229960.
[6] Wu JD, Fang HH, Zhu JY, et al. Multi-rater Prism: Learning self-calibrated medical image segmentation from multiple raters. *Sci*

Bull, 2024,69(18):2906-2919.
[7] Wu Y, Xie XJ, Zhu JH, et al. Overview and prospects of DNA sequence visualization. *Int J Mol Sci*, 2025,26(2):477.
[8] Kharroubi AT, Darwish HM. Diabetes mellitus: The epidemic of the century. *World J Diabetes*, 2015,6(6):850-867.
[9] Magliano DJ, Boyko EJ, IDF Diabetes Atlas Committee. *IDF Diabetes Atlas*. Brussels: International Diabetes Federation, 2021.
[10] Bawankar P, Shanbhag N, S Smitha K, et al. Sensitivity and specificity of automated analysis of single-field non-mydratic fundus photographs by Bosch DR Algorithm-Comparison with mydratic fundus photography (ETDRS) for screening in undiagnosed diabetic retinopathy. *PLoS One*, 2017,12(12):e0189854.
[11] 黄海燕,李德爽,谷浩,等. OCT 和 OCTA 生物学标志物在糖尿病性黄斑水肿预后和监测中的应用. *国际眼科杂志*, 2024,24(5):743-748.
[12] Cicinelli MV, Rabiolo A, Sacconi R, et al. Optical coherence tomography angiography in dry age-related macular degeneration. *Surv Ophthalmol*, 2018,63(2):236-244.
[13] Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs *via* deep learning. *Nat Biomed Eng*, 2018,2(3):158-164.
[14] Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 2016,316(22):2402-2410.
[15] Gulshan V, Rajan RP, Widner K, et al. Performance of a deep-learning algorithm vs manual grading for detecting diabetic retinopathy in India. *JAMA Ophthalmol*, 2019,137(9):987-993.
[16] Gu XW, Zhou Y, Zhao JC, et al. Diagnostic performance and generalizability of deep learning for multiple retinal diseases using bimodal imaging of fundus photography and optical coherence tomography. *Front Cell Dev Biol*, 2025,13:1665173.
[17] Tan TE, Ng YP, Calhoun C, et al. Detection of center-involved diabetic macular edema with visual impairment using multimodal artificial intelligence algorithms. *Ophthalmol Retina*, 2025,9(10):955-963.
[18] Shahriari MH, Sabbaghi H, Asadi F, et al. Artificial intelligence in screening, diagnosis, and classification of diabetic macular edema: a systematic review. *Surv Ophthalmol*, 2023,68(1):42-53.
[19] Shen YJ, Chen Q, He XY, et al. Automated report generation in ophthalmology: integrating artificial intelligence, multimodal imaging, and clinical data. *Ophthalmol Ther*, 2026,15(3):1001-1020.
[20] Fleckenstein M, Keenan TDL, Guymer RH, et al. Age-related macular degeneration. *Nat Rev Dis Primers*, 2021,7:31.
[21] 张超,祁媛媛,王树纲,等. 人工智能在黄斑疾病中的应用及研究进展. *国际眼科杂志*, 2025,25(7):1094-1098.
[22] Li F, Chen H, Liu Z, et al. Fully automated detection of retinal disorders by image-based deep learning. *Graefes Arch Clin Exp Ophthalmol*, 2019,257(3):495-505.
[23] Khalid S, Akram MU, Hassan T, et al. Automated segmentation and quantification of drusen in fundus and optical coherence tomography images for detection of ARMD. *J Digit Imaging*, 2018,31(4):464-476.
[24] Yoo TK, Choi JY, Seo JG, et al. 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. *Med Biol Eng Comput*, 2019,57(3):677-687.
[25] Biswas S, Lin C, Leung CK. Evaluation of a myopic normative database for analysis of retinal nerve fiber layer thickness. *JAMA Ophthalmol*, 2016,134(9):1032-1039.
[26] Cheung CY, Chan N, Leung CK. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: impact of signal strength on analysis of the RNFL map. *Asia Pac J Ophthalmol*, 2012,1(1):19-23.
[27] Cen LP, Ji J, Lin JW, et al. Automatic detection of 39 fundus

diseases and conditions in retinal photographs using deep neural networks. *Nat Commun*, 2021,12(1):4828.

[28] Shi M, Lokhande A, Fazli MS, et al. Artifact-tolerant clustering-guided contrastive embedding learning for ophthalmic images in glaucoma. *IEEE J Biomed Health Inform*, 2023,27(9):4329-4340.

[29] Mariotoni EB, Datta S, Shigueoka LS, et al. Deep learning-assisted detection of glaucoma progression in spectral-domain OCT. *Ophthalmol Glaucoma*, 2023,6(3):228-238.

[30] Medeiros FA, Jammal AA, Mariotoni EB. Detection of progressive glaucomatous optic nerve damage on fundus photographs with deep learning. *Ophthalmology*, 2021,128(3):383-392.

[31] Mariotoni EB, Jammal AA, Urata CN, et al. Quantification of retinal nerve fibre layer thickness on optical coherence tomography with a deep learning segmentation-free approach. *Sci Rep*, 2020,10(1):402.

[32] Gutierrez A, Chen TC. Artificial intelligence in glaucoma: posterior segment optical coherence tomography. *Curr Opin Ophthalmol*, 2023,34(3):245-254.

[33] Asaoka R, Murata H, Hirasawa K, et al. Using deep learning and transfer learning to accurately diagnose early-onset glaucoma from macular optical coherence tomography images. *Am J Ophthalmol*, 2019,198:136-145.

[34] Shibata N, Tanito M, Mitsuhashi K, et al. Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. *Sci Rep*, 2018,8(1):14665.

[35] Hwang EE, Chen DK, Han Y, et al. Utilization of image-based deep learning in multimodal glaucoma detection neural network from a primary patient cohort. *Ophthalmol Sci*, 2025,5(3):100703.

[36] Chuter B, Joshi V, Hallaj S, et al. A novel multimodal implementation of a foundation artificial intelligence model using optic nerve head fundus photographs and OCT imaging for glaucoma detection. *Ophthalmol Sci*, 2026,6(2):101012.

[37] 李涛, 周晓东. 高度近视眼底形态特征的研究进展. *中国眼耳鼻喉科杂志*, 2018,18(6):434-437.

[38] 潘新蕾. 人工智能基于眼底彩照与 OCT 的病理性近视 ATN 分级研究. *北京协和医学院*, 2022.

[39] Xu ZY, Yang YJ, Chen H, et al. Enhancing pathological myopia diagnosis: a bimodal artificial intelligence approach integrating fundus

and optical coherence tomography imaging for precise atrophy, traction and neovascularisation grading. *Br J Ophthalmol*, 2025,109(10):1179-1186.

[40] Choi EY, Kim D, Kim J, et al. Predicting branch retinal vein occlusion development using multimodal deep learning and pre-onset fundus hemisection images. *Sci Rep*, 2025,15(1):2729.

[41] Barbosa G, Carvalho E, Guerra A, et al. Deep learning to distinguish edema secondary to retinal vein occlusion and diabetic macular edema: a multimodal approach using OCT and infrared imaging. *J Clin Med*, 2025,14(3):1008.

[42] Wang M, Zhang X, Li D, et al. The potential of artificial intelligence reading label system on the training of ophthalmologists in retinal diseases, a multicenter bimodal multi-disease study. *BMC Med Educ*, 2025,25(1):503.

[43] Xu J, Shen J, Jiang Q, et al. A multi-modal fundus image based auxiliary location method of lesion boundary for guiding the layout of laser spot in central serous chorioretinopathy therapy. *Comput Biol Med*, 2023,155:106648.

[44] Yoo TK, Kim SH, Kim M, et al. DeepPDT-Net: predicting the outcome of photodynamic therapy for chronic central serous chorioretinopathy using two-stage multimodal transfer learning. *Sci Rep*, 2022,12(1):18689.

[45] Sharma S, Gupta S, Gupta D, et al. HTLML: hybrid AI based model for detection of Alzheimer's disease. *Diagnostics (Basel)*, 2022,12(8):1833.

[46] Ueda E, Watanabe M, Nakamura D, et al. Distinct retinal reflectance spectra from retinal hyperspectral imaging in Parkinson's disease. *J Neurol Sci*, 2024,461:123061.

[47] Tan L, Liu Y, Liu J, et al. Association between insulin resistance and uncontrolled hypertension and arterial stiffness among US adults: a population-based study. *Cardiovasc Diabetol*, 2023,22(1):311.

[48] Jiang JK, Liu JY, Yang J, et al. Optical coherence tomography evaluation of choroidal structure changes in diabetic retinopathy patients: a systematic review and meta-analysis. *Front Med*, 2022,9:986209.

[49] Pellegrini C, Ravaioli F, De Fanti S, et al. Detection of Brain-Derived Cell-Free DNA in Plasma. *Diagnostics (Basel)*, 2024,14(22):2541.