

多类视网膜疾病OCT图像分类方法研究

徐国荣

上海理工大学健康科学与工程学院, 上海

收稿日期: 2024年12月26日; 录用日期: 2025年1月18日; 发布日期: 2025年1月28日

摘要

视网膜疾病是当前威胁人类视觉健康的重要问题之一,其早期诊断和干预对预防视力损害具有重要意义。光学相干断层扫描作为一种无创成像技术,在视网膜疾病诊断中发挥着关键作用。本文提出了一种基于改进DenseNet的深度学习模型,用于多类视网膜OCT图像的自动分类。在OCT-C8数据集上进行实验,结果表明改进后的模型在八类视网膜疾病的分类任务中表现优异,平均准确率达到99.41%。与现有其他方法相比,本文提出的模型展现出更优的分类性能。

关键词

视网膜疾病, 光学相干断层扫描, 深度学习

Research on OCT Image Classification Methods for Multiple Types of Retinal Diseases

Guorong Xu

School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Dec. 26th, 2024; accepted: Jan. 18th, 2025; published: Jan. 28th, 2025

Abstract

Retinal diseases are one of the most important problems threatening human visual health, and their early diagnosis and intervention are of great significance in preventing visual impairment. Optical coherence tomography, as a non-invasive imaging technique, plays a key role in the diagnosis of retinal diseases. In this paper, a deep learning model based on improved DenseNet is proposed for the automatic classification of multi-class retinal OCT images. Experiments on the OCT-C8 dataset show that the improved model performs well in the classification task of eight retinal diseases with

an average accuracy of 99.41%. Compared with other existing methods, the proposed model shows better classification performance.

Keywords

Retinal Diseases, Optical Coherence Tomography, Deep Learning

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

视网膜相关疾病已成为当前威胁全球视觉健康的重要问题，不仅显著降低患者的生活质量和工作能力，还严重影响其社会参与度[1]。在众多视网膜疾病中，糖尿病性视网膜病变(DR)和年龄相关性黄斑变性(AMD)最为常见。研究显示，DR 在单眼引起的视力损害比例高达 32%，双眼达到 60% [2]。而 AMD 作为 50 岁以上人群致盲的主要原因之一，预计到 2040 年全球患病人数将达到 2.88 亿。美国国家健康和营养调查(NHANES)的分析显示，超过 70%的 DR 患者和 AMD 患者并不了解自身病情发展情况[3]。

光学相干断层扫描(Optical Coherence Tomography, OCT)是近年来广泛应用于眼科领域的非侵入性成像技术，能够提供视网膜结构的高分辨率横断面图像，并准确测定层间及总体厚度。这种能力对于及时识别视网膜病理改变具有关键作用，已成为评估多种眼病的重要工具[4]。

OCT 技术为视网膜疾病诊断提供了大量的高质量影像学数据，在此基础上，越来越多的研究人员开始探索和开发自动化技术，以提高 OCT 图像的分析的准确性。Wang 等人提出使用具有特征重用功能的 CliqueNet 网络结构，在公开的 OCT 数据集上实现了年龄相关性黄斑变性、糖尿病性黄斑水肿和正常样本的三分类任务，虽然取得了 98%以上的准确率，但仅局限于三种类别的分类识别[5]。Subramanian 等人使用 VGG16 网络结构实现了 OCT 图像的 8 类视网膜疾病分类任务，在 OCT-C8 数据集上取得了 97% 的准确率，但模型参数量大且训练时间长[6]。本研究基于深度学习技术，构建了一个针对 8 类视网膜疾病的 OCT 图像分类的神经网络。通过对 24,000 张 OCT 图像的训练和验证，实现了对多类视网膜疾病的高精度识别。

2. 数据集与实验方法

2.1. 数据集

在本研究中，使用了来自 Subramanian 等人整理并发布在 Kaggle 平台的公开数据集“Retinal OCT Image Classification-C8” [6]。该数据集包含八种不同类型的视网膜状况，年龄相关性黄斑变性(Age-related Macular Degeneration, AMD)、脉络膜新生血管(Choroidal Neovascularization, CNV)、中心性浆液性视网膜病变(Central Serous Retinopathy, CSR)、糖尿病性黄斑水肿(Diabetic Macular Edema, DME)、糖尿病视网膜病变(Diabetic Retinopathy, DR)、玻璃膜疣(Drusen)、黄斑裂孔(Macular Hole, MH)以及正常眼底(Normal)。图 1 中的样本图像来源于 OCT-C8 数据集，分别从 8 种视网膜状况的训练集中随机选取，每个类别选取一张，最终组合而成。这些样本清晰地体现了不同视网膜状况在 OCT 成像中的特征，为后续分析提供了直观的参考。

该数据集的构建融合了多个来源，其中 AMD、CNV、DME、DRUSEN 和正常眼底图像采集自 Kaggle

平台，而 CSR、DR 和 MH 则来自开源医学影像库。为确保数据质量和类别平衡性，创建者对原始图像进行了标准化处理，并对各类别的样本数量进行了均衡，如表 1 所示，数据集总共包含 24,000 张高质量 OCT 图像，按照约 77:12:12 的比例划分为训练集、验证集和测试集，每个类别在各个集合中的样本分布均保持一致，这种的数据划分有助于模型的训练和性能评估。

Table 1. OCT-C8 dataset details
表 1. OCT-C8 数据集详细信息

类别	英文缩写	训练集数量	验证集数量	测试集数量	总数
年龄相关性黄斑变性	AMD	2300	350	350	3000
脉络膜新生血管	CNV	2300	350	350	3000
中心性浆液性视网膜病变	CSR	2300	350	350	3000
糖尿病性黄斑水肿	DME	2300	350	350	3000
糖尿病视网膜病变	DR	2300	350	350	3000
玻璃膜疣	Drusen	2300	350	350	3000
黄斑裂孔	MH	2300	350	350	3000
正常眼底	Normal	2300	350	350	3000

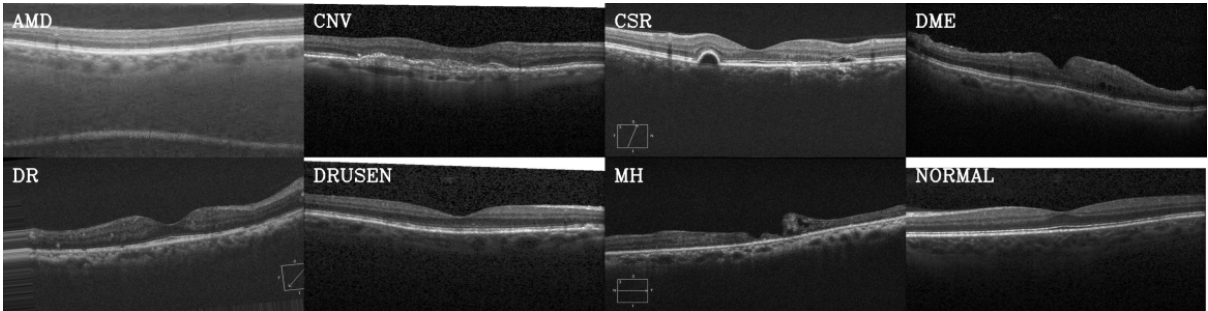


Figure 1. Sample display of eight different retinal conditions in the OCT-C8 dataset
图 1. OCT-C8 数据集中八种不同视网膜状况的样本展示

2.2. 整体架构

本研究对 DenseNet 网络结构进行了多方面的改进和优化，以更好地适应 OCT 图像分类任务的特点。首先，针对 OCT 医学影像的固有特征，将输入层由传统的三通道 RGB 结构调整为单通道灰度图像输入。其次，为了提升模型的特征提取能力，在 DenseBlock 中集成了高效通道注意力机制(Efficient Channel Attention, ECA)模块，该机制通过轻量级的一维卷积操作自适应地学习通道间的关联性[7]。

为了提升分类性能，对 DenseNet 网络中的单层线性分类器进行了改进，设计了一种更复杂的多层分类器结构。新的分类器由两层全连接层组成，第一层全连接层将输入特征映射至 512 维中间表示，通过 ReLU 激活函数引入非线性特性，随后经过批量归一化层对特征进行归一化处理，提升训练过程中的数据分布稳定性。接着，通过 Dropout 正则化(丢弃率为 0.2)进一步抑制过拟合，最后的全连接层将特征映射至 8 类输出概率。

在模型训练方面，采用了改进的参数初始化策略。对于卷积层，根据卷积核大小的不同采用不同的初始化方法，对于大于 1×1 的卷积核使用 Kaiming 初始化，这种方法考虑了网络的非线性特性，有助于梯度在深层网络中的稳定传递[8]。对于 1×1 的卷积核则采用标准正态分布初始化。本研究保留了

DenseNet 网络的核心特点，包括稠密连接结构和特征重用机制。在此基础上，引入了动态内存管理机制，在训练过程中释放不必要的中间特征图内存，降低资源占用。这些改进不仅增强了训练过程的稳定性，同时保证了较高的分类准确率。网络的整体结构图如图 2 所示。

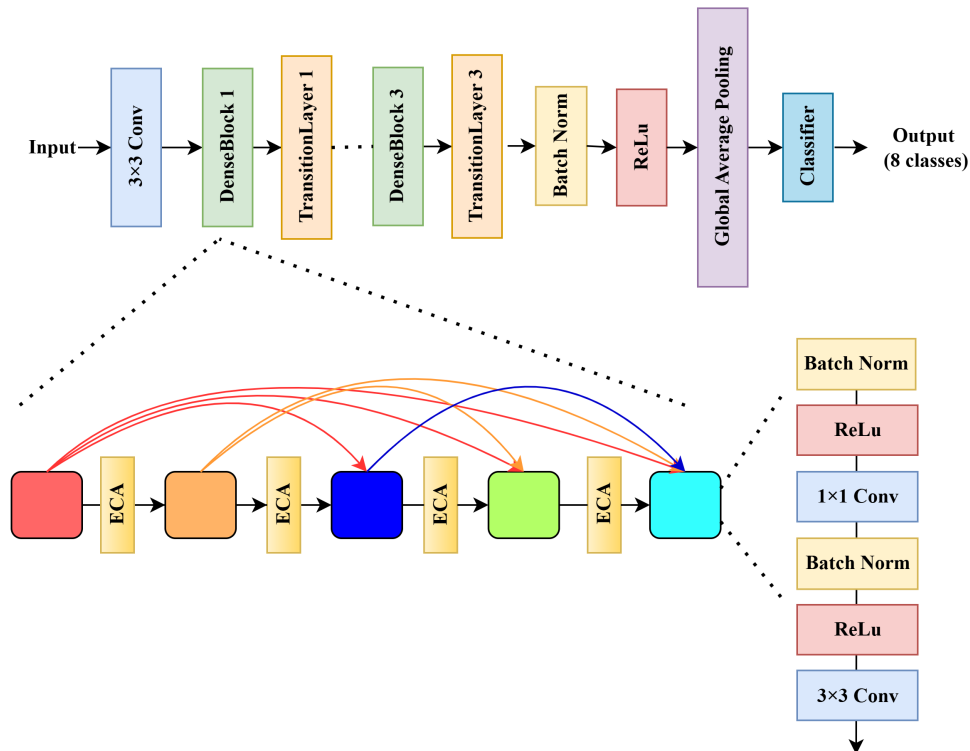


Figure 2. The structure diagram of the improved DenseNet
图 2. 改进后的 DenseNet 结构图

2.3. 高效通道注意力机制 ECA

ECA 是一种轻量化的通道注意力模块，其核心思想是在保持低计算复杂度的同时，有效建模通道间的相互依赖关系。在 ECA 模块中，首先对输入特征图进行全局平均池化操作，将每个通道的空间信息压缩为单一数值，实现通道信息的全局聚合。随后，通过一维卷积层处理池化后的特征，实现局部通道间的交互。一维卷积的核大小(k)由映射关系 $\psi(C)$ 确定，其中 C 是输入通道数，映射关系由公式(1)给出：

$$k = \psi(C) = \left\lfloor \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}} \quad (1)$$

这里， γ 和 b 是超参数，通常设置为 2 和 1， $\lfloor \cdot \rfloor_{\text{odd}}$ 表示结果取最近的奇数，确保核大小为奇数，有助于保持通道间的对称性。一维卷积层的输出随后通过 sigmoid 激活函数，这个函数将输出值压缩到 0 和 1 之间，生成每个通道的权重系数。sigmoid 函数的输出与原始输入特征图进行元素级乘法操作，这样每个通道的输出特征图都被其对应的权重系数所调整，从而实现通道间的重新校准。最后，经过元素级乘法，得到的特征图作为 ECA 模块的输出，这些输出特征图包含了经过通道注意力机制增强的信息。

与传统注意力模块不同，ECA 模块避免了维度压缩操作，保持了通道与权重之间的直接对应关系。同时，通过局部通道交互策略和自适应核大小选择，使得 ECA 模块在多种视觉任务中展现出了较好的性能提升[9]。ECA 模块的结构图如图 3 所示。

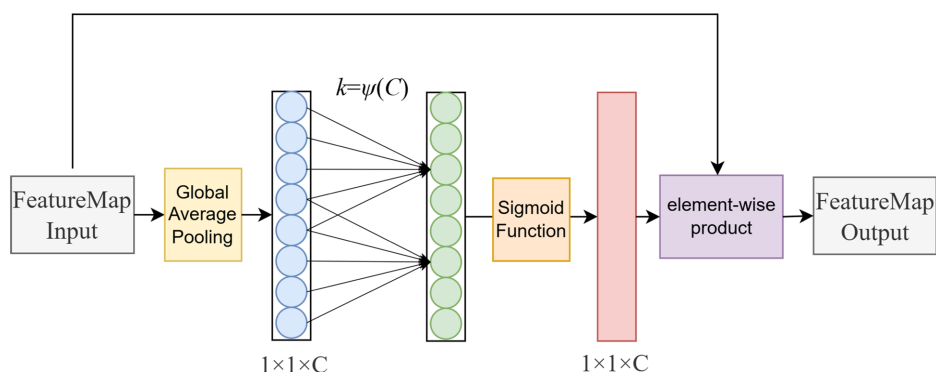


Figure 3. Efficient channel attention
图 3. 高效通道注意力机制

3. 实验流程

3.1. 实验环境与参数设置

本实验在基于 Ubuntu 20.04 操作系统的深度学习环境下进行, 采用 NVIDIA RTX 4090D (24GB) GPU 进行模型训练和推理。系统配置包括 AMD EPYC 9754 处理器(18vCPU)。软件环境采用 Python 3.8 与 PyTorch 1.11.0 深度学习框架, 并基于 CUDA 11.3 实现 GPU 加速。在数据预处理阶段, 将输入图像统一调整为 224×224 大小, 并采用数据增强策略以提升模型泛化能力。具体的数据增强方法包括随机水平翻转、 $\pm 10^\circ$ 随机旋转、 $\pm 5\%$ 随机平移以及 $\pm 20\%$ 的亮度和对比度调整。此外, 对输入数据进行标准化处理, 设置均值为 0.5, 标准差为 0.5。

训练中, 采用 batch size 为 32 的小批量随机梯度下降方法。选用 AdamW 优化器, 初始学习率设置为 0.001, 权重衰减系数为 0.01。为了动态调整学习率, 引入根据指标调整学习率的调度策略, 当验证集性能在连续 5 个 epoch 内没有改善时, 学习率降低为原来的 0.1 倍。训练过程设置最大训练轮次为 30, 为防止梯度爆炸, 设置梯度裁剪阈值为 1.0。选用交叉熵损失函数(Cross Entropy Loss)作为模型的优化目标。为提高训练效率, 数据加载采用 8 线程并行处理。

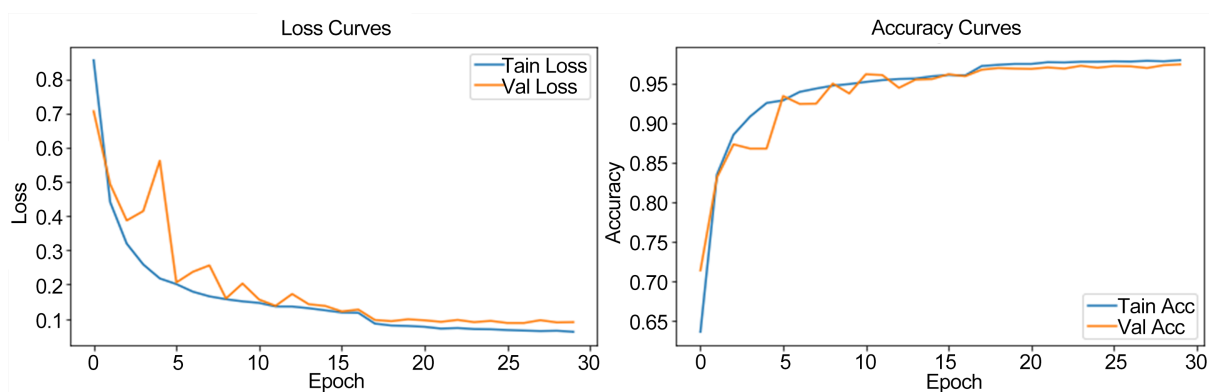


Figure 4. Loss function and accuracy variation curve
图 4. 损失函数与准确率变化曲线

3.2. 训练过程

如图 4 所示的训练过程中, 模型在前 5 个 epoch 表现出快速的学习能力, 损失值从 0.85 迅速下降至

0.3 左右。随后训练趋于平稳，在第 30 个 epoch 时，训练集和验证集的损失值均稳定在 0.1 附近。准确率方面，模型在训练后期达到了 95% 以上的水平，且训练集和验证集的准确率曲线保持接近，表明模型具有良好的泛化能力，没有出现拟合。

3.3. 评价指标

为了全面评估模型在 OCT 图像多分类任务中的性能，本文采用了准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F1 分数(F1-score)作为评价指标。其中：

准确率反映了模型整体的分类准确程度，其计算公式(2)为：

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

精确率表示在所有预测为正类的样本中真实为正类的比例，其计算公式(3)为：

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

召回率衡量了模型检测出正确样本的能力，其计算公式(4)为：

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

F1 分数是精确率和召回率的调和平均值，其计算公式(5)为：

$$F1-Score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (5)$$

TP 表示模型正确将病例归为阳性的数量， TN 表示模型正确识别为阴性的数量。 FP 代表模型错误地将阴性样本判断为阳性的数量，而 FN 则表示模型将阳性病例误判为阴性的数量。

4. 实验结果与分析

4.1. 分类结果的可视化

为了全面评估模型的分类性能，采用混淆矩阵和准确率柱状图两种可视化方法来展示实验结果。图 5 展示了基于 OCT-C8 数据集的混淆矩阵，直观地反映了模型在各类别间的分类表现。从混淆矩阵中可以观察到，模型在大多数类别上都表现出极高的分类准确性。AMD、CSR、DR 和 MH 这四个类别都达到了最佳的分类效果，预测数量均为 350 例，没有出现任何错误分类的情况。这表明模型能够准确捕捉这些疾病类型的特征性表现。然而，在其他类别中也出现了一些误分类现象。CNV 类别中有 322 例被正确分类，但有 10 例被误判为 DME，16 例被误判为 DRUSEN，2 例被误判为 NORMAL，这反映出 CNV 与这些类别可能存在一些相似的图像特征。DME 类别显示了较好的分类性能，336 例被正确分类，仅有 3 例被误判为 CNV，3 例被误判为 DRUSEN，8 例被误判为 NORMAL。DRUSEN 类别的表现也相当不错，333 例被正确分类，仅有少量样本被误分类到其他类别。NORMAL 类别获得了 343 例正确分类，表现出良好的识别能力。

图 6 通过柱状图形式更直观地展示了各类别的分类准确率。从图中可以清楚地看到，模型在 AMD、CSR、DR 和 MH 这四个类别上都达到了 100% 的准确率。NORMAL 类别紧随其后，达到 98.00% 的高准确率。DME 和 DRUSEN 类别也分别达到了 96.00% 和 95.14% 的良好表现。值得注意的是，CNV 类别的准确率相对较低，为 92.00%，这与混淆矩阵中观察到的误分类情况相对应。图中的红色虚线表示总体准确率，达到了 97.64% 的高水平，这些可视化结果表明了模型的可靠性。

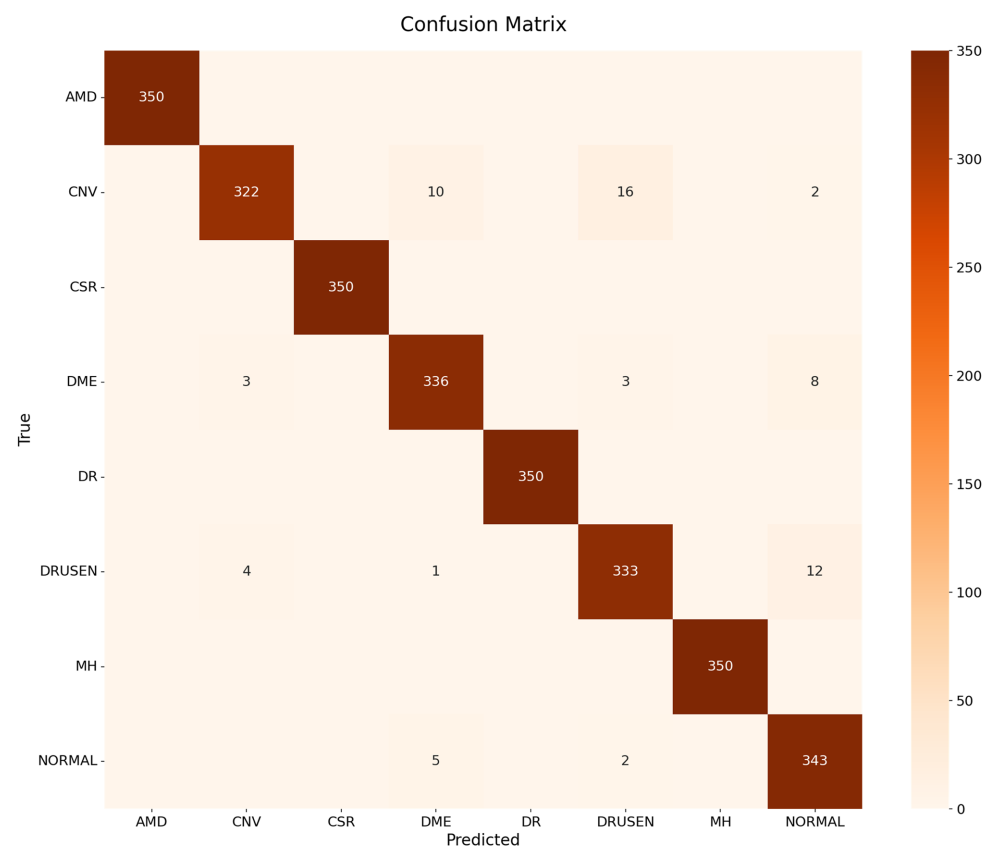


Figure 5. Confusion matrix based on OCT-C8
图 5. 基于 OCT-C8 的混淆矩阵

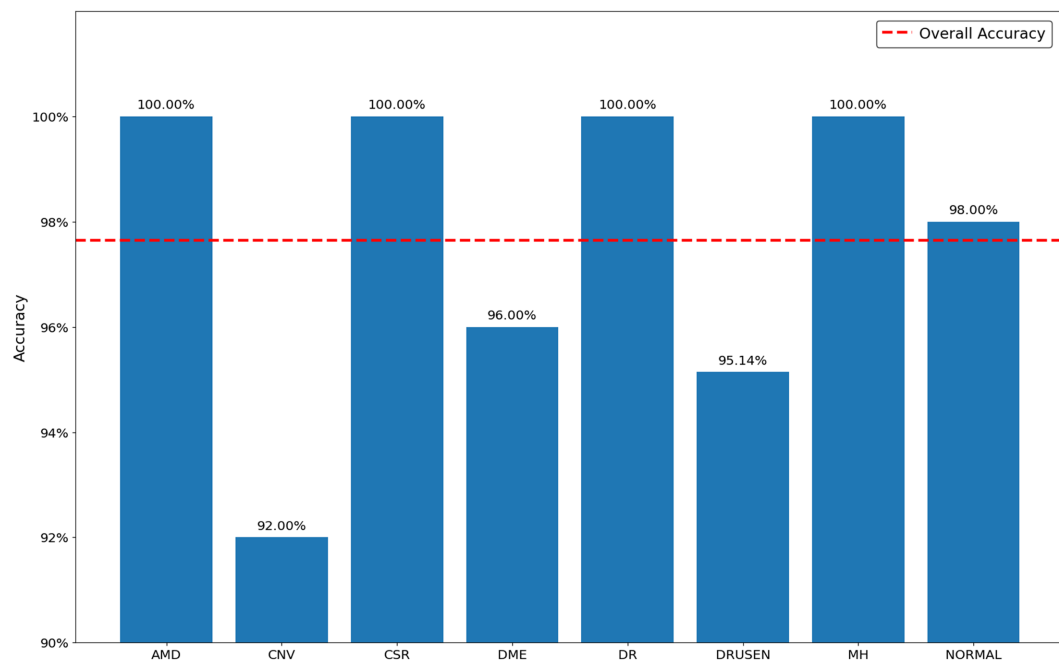


Figure 6. Model accuracy performance on different retinal conditions
图 6. 模型在不同视网膜状况上的准确率表现

4.2. 实验结果的评价指标

表 2 展示了 DenseNet 和改进后的 Improved DenseNet 在各个类别上的详细性能表现。通过对比分析可以发现, Improved DenseNet 在大多数评价指标上都取得了一定提升。特别是 DRUSEN 类别的改进效果, 其召回率从 86.29%提升至 95.14%, F1-Score 相应提高至 94.60%, 表明改进策略有效提升了模型对该类别的识别能力。

Table 2. Performance comparison between DenseNet and improved models

表 2. DenseNet 与改进模型的性能对比

模型	类别	Accuracy	Precision	Recall	F1-Score
DenseNet	AMD	99.93%	100%	99.43%	99.71%
	CNV	98.32%	89.76%	97.71%	93.57%
	CSR	99.93%	99.43%	100%	99.72%
	DME	98.71%	95.64%	94.00%	94.81%
	DR	100%	100%	100%	100%
	DRUSEN	98.04%	97.73%	86.29%	91.65%
	MH	100%	100%	100%	100%
	NORMAL	98.71%	92.90%	97.14%	94.97%
Improved DenseNet	AMD	100%	100%	100%	100%
	CNV	98.75%	97.87%	92.00%	94.85%
	CSR	100%	100%	100%	100%
	DME	98.93%	95.45%	96.00%	95.73%
	DR	100%	100%	100%	100%
	DRUSEN	98.64%	94.07%	95.14%	94.60%
	MH	100%	100%	100%	100%
	NORMAL	98.96%	93.97%	98.00%	98.70%

4.3. 模型性能对比分析

为了进一步验证本文提出的改进模型的有效性, 与其他视网膜疾病分类模型进行了对比, 表 3 展示了不同模型在相同数据集上的性能比较结果。从实验结果可以看出, 本文提出的 Improved DenseNet 模型在各项评价指标上都取得了最优表现。具体而言, 改进模型获得了 99.41%的平均准确率(mAccuracy), 相比原始 DenseNet 的 99.21%有所提升, 同时显著优于其他对比模型。在平均精确率(mPrecision)和平均召回率(mRecall)方面, 改进模型分别达到了 97.67%和 97.64%, 展现出稳定和均衡的分类性能。特别是在综合评价指标平均 F1-Score(mF1-Score)上, 改进模型达到了 97.99%的高水平, 相比原始 DenseNet 的 96.80%有了明显提升, 也领先于其他模型。

相比之下, He 等人[10]提出的方法在各项指标上均达到了 97%左右的水平, 而 Subramanian 等人[6]的模型虽然取得了 99.30%的较高准确率, 但在其他指标上略显不足。Karthik 等人[11]的模型各项指标则相对较低, 在 92%~93%之间。这些对比结果充分证明了本文提出的改进策略的有效性, 不仅提升了模型的整体性能, 还在各项指标上都实现了更加平衡的表现。

Table 3. Comparison with other advanced models
表 3. 与其他先进模型的对比

模型	mAccuracy	mPrecision	mRecall	mF1-Score
Improved DenseNet	99.41%	97.67%	97.64%	97.99%
DenseNet	99.21%	96.93%	96.82%	96.80%
He <i>et al.</i> [10]	97.12%	97.13%	97.13%	97.10%
Subramanian <i>et al.</i> [6]	99.30%	97.25%	97.13%	97.25%
Karthik <i>et al.</i> [11]	92.40%	93.00%	92.00%	92.00%

5. 总结与展望

本研究提出了一种改进的 DenseNet 深度学习模型，用于多类视网膜 OCT 图像的自动分类。通过引入高效通道注意力机制、优化分类器结构和修改参数初始化等多项改进策略，模型的性能得到了整体提升。改进后的模型在 AMD、CSR、DR 和 MH 四类疾病的分类中均达到 100% 的准确率，在整个 OCT-C8 数据集上取得了 99.41% 的平均准确率和 97.99% 的平均 F1-Score。模型对此前较难识别的 DRUSEN 类别也表现出了显著的改善，其召回率从 86.29% 提升至 95.14%，充分验证了改进策略的有效性。实验结果表明，该模型已经具备了稳定识别八种不同视网膜状况的能力，在分类性能和泛化能力上均达到了较高水平，可以为视网膜疾病的临床辅助诊断提供有效的技术支持。

虽然本研究取得了良好的成果，但在 CNV 类别的分类中准确率相对较低，并且存在少量与 DME 和 DRUSEN 的误分类现象。这表明，模型在处理具有相似特征的疾病时仍需进一步优化，同时还需在更大规模和更多样化的数据集上验证其泛化能力。此外，如何将模型与临床实践更紧密地结合，也是未来研究的重要方向。

参考文献

- [1] Zhao, X., Lin, J., Yu, S., Xie, L., Jin, L., Xiong, K., *et al.* (2022) Incidence, Causes and Risk Factors of Vision Loss in Rural Southern China: 6-Year Follow-Up of the Yangxi Eye Study. *British Journal of Ophthalmology*, **107**, 1190-1196. <https://doi.org/10.1136/bjophthalmol-2021-320892>
- [2] Zhou, C., Li, S., Ye, L., Chen, C., Liu, S., Yang, H., *et al.* (2023) Visual Impairment and Blindness Caused by Retinal Diseases: A Nationwide Register-Based Study. *Journal of Global Health*, **13**, Article No. 04126. <https://doi.org/10.7189/jogh.13.04126>
- [3] Gibson, D.M. (2012) Diabetic Retinopathy and Age-Related Macular Degeneration in the U.S. *American Journal of Preventive Medicine*, **43**, 48-54. <https://doi.org/10.1016/j.amepre.2012.02.028>
- [4] Elsharkawy, M., Elrazzaz, M., Ghazal, M., Alhalabi, M., Soliman, A., Mahmoud, A., *et al.* (2021) Role of Optical Coherence Tomography Imaging in Predicting Progression of Age-Related Macular Disease: A Survey. *Diagnostics*, **11**, Article No. 2313. <https://doi.org/10.3390/diagnostics11122313>
- [5] Wang, D. and Wang, L. (2019) On OCT Image Classification via Deep Learning. *IEEE Photonics Journal*, **11**, 1-14. <https://doi.org/10.1109/jphot.2019.2934484>
- [6] Subramanian, M., Shanmugavadivel, K., Naren, O.S., Premkumar, K. and Rankish, K. (2022) Classification of Retinal OCT Images Using Deep Learning. 2022 *International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, 25-27 January 2022, 1-7. <https://doi.org/10.1109/iccci54379.2022.9740985>
- [7] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W. and Hu, Q. (2020) ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 11534-11542. <https://doi.org/10.1109/cvpr42600.2020.01155>
- [8] He, K., Zhang, X., Ren, S. and Sun, J. (2015) Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. 2015 *IEEE International Conference on Computer Vision (ICCV)*, Santiago, 7-13 December 2015, 1026-1034. <https://doi.org/10.1109/iccv.2015.123>

-
- [9] Lou, Q., Li, Y., Qian, Y., Lu, F. and Ma, J. (2022) Mammogram Classification Based on a Novel Convolutional Neural Network with Efficient Channel Attention. *Computers in Biology and Medicine*, **150**, Article ID: 106082. <https://doi.org/10.1016/j.compbiomed.2022.106082>
 - [10] He, J., Wang, J., Han, Z., Ma, J., Wang, C. and Qi, M. (2023) An Interpretable Transformer Network for the Retinal Disease Classification Using Optical Coherence Tomography. *Scientific Reports*, **13**, Article No. 3637. <https://doi.org/10.1038/s41598-023-30853-z>
 - [11] Karthik, K. and Mahadevappa, M. (2023) Convolution Neural Networks for Optical Coherence Tomography (OCT) Image Classification. *Biomedical Signal Processing and Control*, **79**, Article ID: 104176. <https://doi.org/10.1016/j.bspc.2022.104176>