

改进MobileViT模型的脑胶质瘤良恶分类方法

赵菁菁, 郑钟月, 张依林, 张丽艳*

大连交通大学轨道智能工程学院电子与通信工程系, 辽宁 大连

收稿日期: 2026年6月29日; 录用日期: 2026年7月21日; 发布日期: 2026年7月30日

摘要

脑胶质瘤良恶性分级对术前个体化诊疗方案的制定具有重要临床价值。磁共振成像(MRI)是胶质瘤无创诊断的首选工具, 但人工阅片受主观因素影响, 诊断一致性难以保证。针对现有深度学习方法存在的感受野受限、通道特征自适应建模不足及小样本泛化能力弱等问题, 本文提出一种基于改进MobileViT的脑胶质瘤良恶性分类方法。在MobileViT骨干网络基础上, 引入空洞卷积金字塔模块(Efficient Spatial Pyramid module, ESP模块)构建多膨胀率并行分支与层级特征融合结构, 增强对病灶多尺度上下文特征的捕获能力; 引入通道注意力模块(Channel Attention module, CA模块)自适应生成通道注意力权重, 强化判别性通道特征响应; 并结合学习率线性衰减、标签平滑及医学图像迁移学习三种策略优化训练过程。在BraTS2019数据集上的病例级五折交叉验证结果表明, 本文方法ACC达98.87%, SEN、SPE、PPV、NPV分别为98.21%、99.53%、99.52%及98.23%, 模型参数量仅7.6 M, 综合性能优于MobileNetV2、ResNet101、EfficientNetB3等经典网络及现有同类方法。

关键词

脑胶质瘤, 良恶性分级, MobileViT, 空洞卷积, 通道注意力

A Glioma Malignancy Grading Method Based on an Improved MobileViT Model

Jingjing Zhao, Zhongyue Zheng, Yilin Zhang, Liyan Zhang*

Department of Electronic Communications, School of Railway Intelligent Engineering, Dalian Jiaotong University, Dalian Liaoning

Received: June 29, 2026; accepted: July 21, 2026; published: July 30, 2026

Abstract

Accurate preoperative malignancy grading of glioma is essential for individualized treatment plan-

*通讯作者。

文章引用: 赵菁菁, 郑钟月, 张依林, 张丽艳. 改进 MobileViT 模型的脑胶质瘤良恶分类方法[J]. 图像与信号处理, 2026, 15(3): 415-425. DOI: 10.12677/jisp.2026.153037

ning. Although magnetic resonance imaging (MRI) is the preferred modality for non-invasive glioma diagnosis, manual interpretation is susceptible to subjective bias and inconsistency. To address the limitations of existing deep learning methods, including constrained receptive fields, insufficient channel feature modeling, and poor small-sample generalization, this paper proposes an improved MobileViT-based glioma grading method. An Efficient Spatial Pyramid (ESP) module is introduced for multi-scale contextual feature extraction, and a Channel Attention (CA) module is incorporated to reinforce discriminative channel responses. Three training strategies—linear learning rate decay, label smoothing, and medical image-based transfer learning—are further applied. Five-fold cross-validation on BraTS2019 yields an accuracy (ACC) of 98.87%, with sensitivity (SEN), specificity (SPE), positive predictive value (PPV), and negative predictive value (NPV) of 98.21%, 99.53%, 99.52%, and 98.23%, respectively. With only 7.6 M parameters, the proposed method outperforms MobileNetV2, ResNet101, and EfficientNetB3, achieving a favorable balance between lightweight design and classification performance.

Keywords

Glioma Malignancy Grading, MobileViT, Dilated Convolution, Channel Attention

Copyright © 2026 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

脑胶质瘤是中枢神经系统最常见的原发性颅内肿瘤，起源于脑组织中的神经胶质细胞，占成人恶性原发性脑肿瘤的 75% [1]。依据世界卫生组织(WHO)病理分级标准，胶质瘤被划分为I至IV级，其中I、II级为低级别胶质瘤(LGG)，临床上通常归为良性；III、IV级为高级别胶质瘤(HGG)，即恶性胶质瘤[2]。良性胶质瘤手术切除后预后相对较好；恶性胶质瘤侵袭性强，术后须长期放化疗，整体预后差，长期生存率低[3]。两类肿瘤在治疗策略与预后走向上差异显著，术前准确判断良恶性是制定个体化诊疗方案的重要前提[4]。

磁共振成像(MRI)凭借多序列成像、软组织对比度高、无电离辐射等特点，已成为胶质瘤无创诊断与术前评估的首选影像工具[5]。然而，MRI 影像的解读高度依赖医师的临床经验，阅片结论易受主观判断与疲劳因素影响，诊断一致性难以保证[6]。并且胶质瘤形态高度异质、边界模糊，良恶性病灶在影像表现上存在一定交叠，人工判读难度较大。随着临床影像检查量的持续增长，研发客观、高效的计算机辅助诊断系统已成为该领域的迫切需求[7]。早期计算机辅助诊断方法以放射组学为代表，通过手工提取直方图统计量、几何形态及纹理等特征，结合传统分类器完成良恶性判别[8]。该类方法可重复性较好，但特征工程依赖专家先验知识，流程繁琐，面对多中心数据时泛化能力不足[9]。深度学习的兴起为上述问题提供了新的解决路径，以卷积神经网络(CNN)为代表的方法能够以数据驱动方式自动学习多层次特征表示，推动了医学图像分类研究的快速发展[10]。在基于深度学习的胶质瘤良恶性分类研究中，迁移学习是最常见的技术路线。Yang 等人[11]对预训练 CNN 进行微调，在常规 MRI 上取得了 90.9%的分级准确率；Sajjad 等人[12]引入大规模数据增强策略，实现四级脑肿瘤分类，整体准确率达 90.67%。赵尚义等人[13]在 U-Net 架构中引入多层次特征融合机制，将浅层纹理特征与深层语义特征相结合构建分类系统，分类准确率达 91.67%。此后，注意力机制、多尺度特征融合及残差连接等结构相继被引入，模型的特征表达能力与分类精度进一步提升。然而，现有方法仍存在若干共性局限：标准卷积感受野受限，难以有效

建模病灶的多尺度上下文信息;对特征图各通道的差异性贡献缺乏自适应建模,冗余特征干扰分类决策;主流网络参数量大、推理延迟高,在临床资源受限场景下部署困难。上述问题制约了现有方法在真实临床环境中的推广应用。

针对以上不足,本文提出一种基于改进 MobileViT 的脑胶质瘤良恶性分类方法。主要改进如下:(1) 将空洞卷积金字塔模块(Efficient Spatial Pyramid module, ESP 模块)引入 MobileViT 骨干网络,构建多膨胀率感受野金字塔,增强对胶质瘤病灶多尺度上下文特征的捕获能力;(2) 引入通道注意力模块(Channel Attention module, CA 模块),自适应强化判别性通道特征响应;(3) 提出针对小样本 MRI 分类场景的复合训练优化策略,提升模型的收敛稳定性与泛化性能。所提方法兼顾轻量化与分类精度,旨在为脑胶质瘤术前无创良恶性分级提供高效可靠的智能辅助诊断支撑。

2. ESP-CA-MobileViT 网络

2.1. 网络整体框架

在对特征信息复杂、数据维度高的脑胶质瘤 MRI 图像进行良恶性分类时,不仅需要考虑网络对多尺度病灶特征的提取能力,还需兼顾模型的参数规模与推理效率。MobileViT 将轻量级 MobileNetV2 (MV2) 倒残差模块[14]与 Transformer 全局建模机制有机结合,在保持较低参数数量的同时兼顾局部特征提取与全局上下文建模,适合在医学小样本场景下部署应用。为此,本文选用 MobileViT 作为骨干网络,并针对胶质瘤良恶性分类任务的特点进行改进。

具体改进思路如下:首先,在主干网络的中高层特征提取阶段引入 ESP 模块,通过构建多膨胀率并行分支的金字塔结构,在同一层次上同时捕获不同尺度的上下文语义特征,实现局部纹理特征与大范围语义信息的协同表达;其次,在特征融合阶段引入 CA 模块,增强网络对病灶判别性特征的关注度,降低冗余通道特征对分类决策的干扰;最后,采用学习率线性衰减、标签平滑及迁移学习三种优化策略的组合,提升模型训练效率与泛化能力。图 1 为本文提出模型的网络整体结构。

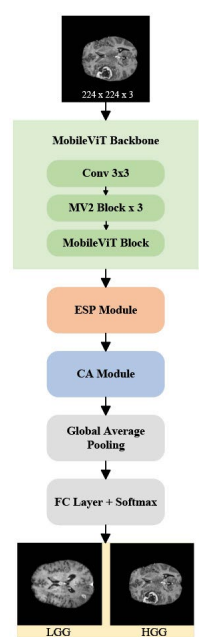


Figure 1. The structure of proposed model
图 1. 模型结构图

2.1.1. 空洞卷积

标准卷积操作的感受野由卷积核尺寸直接决定，若要扩大感受野则须增大卷积核或加深网络层数，前者显著增加参数量，后者则带来更高的计算开销与过拟合风险。空洞卷积[15]通过在卷积核相邻元素之间插入间隔，在不引入额外参数的前提下有效扩大感受野，为多尺度特征提取提供了一种高效的实现途径。设卷积核大小为 K ，膨胀率为 d ，则空洞卷积的等效感受野如公式(1)所示：

$$R_d = K + (K - 1)(d - 1) \quad (1)$$

图 2 为不同膨胀率的空间卷积。以 $K = 3$ 为例，当 $d = 1$ 时等效于普通 3×3 卷积，感受野为 3×3 ；当 $d = 2$ 时感受野扩大至 5×5 ；当 $d = 3$ 时感受野进一步扩大至 7×7 。三种情况下参数量均保持与 3×3 卷积一致，仅为 9 个可学习权重。

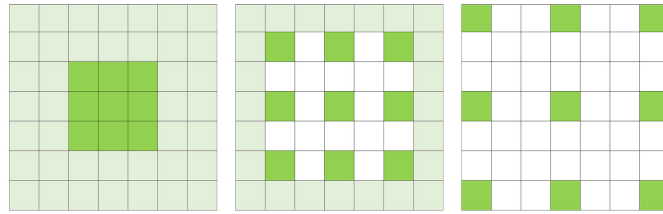


Figure 2. Dilated convolutions with different dilation rates
图 2. 不同膨胀率的空间卷积

对于输入特征图 F ，膨胀率为 d 的空洞卷积输出可表示为公式(2)：

$$y(p) = \sum_{k \in K} w(k) \cdot F(p + d \cdot k) \quad (2)$$

其中 P 为输出特征图上的位置坐标， k 为卷积核内的偏移索引， $w(k)$ 为对应的卷积核权重， d 为膨胀率。当 $d = 1$ 时，式(2)退化为标准卷积。

2.1.2. ESP 模块

脑胶质瘤病灶在 MRI 影像中具有明显的尺度差异性：肿瘤核心区域边界较为清晰，局部纹理特征丰富，而瘤周水肿区域范围较广，与正常组织之间存在模糊过渡。传统单尺度卷积难以同时兼顾局部细节信息与全局上下文语义。针对上述问题，本文在网络中引入 ESP 模块[16]进行多尺度特征提取。该模块利用多分支空洞卷积结构扩展感受野，并结合层级特征融合策略，在控制参数量与计算量的同时，有效增强了模型对不同尺度病灶区域的表征能力。ESP 模块结构如图 3 所示。

设输入特征图为 F ，ESP 模块首先通过 1×1 点卷积对输入进行通道降维，将通道数压缩至 $C/4$ 以降低后续多分支计算的参数量，该过程表示为公式(3)：

$$F' = \text{Conv}_{1 \times 1}(F) \quad (3)$$

降维后的特征图 F' 被送入四路并行分支，本文将四路分支膨胀率设置为 $\{1, 2, 4, 8\}$ ，对应等效感受野分别为 3×3 、 5×5 、 7×7 和 9×9 ，依次覆盖肿瘤核心局部纹理、近周边过渡区域至瘤周水肿大范围上下文，实现对胶质瘤病灶多尺度空间结构的完整建模，第 i 路分支的输出表示为公式(4)：

$$E_i = \text{DConv}_{3 \times 3}^{d=r_i}(F'), \quad i = 1, 2, 3, 4 \quad (4)$$

其中 $r_1 = 1, r_2 = 2, r_3 = 4, r_4 = 8$ ， $\text{DConv}_{3 \times 3}^{d=r_i}(\cdot)$ 表示膨胀率为 r_i 的 3×3 空洞卷积操作。

为抑制较大膨胀率下空洞卷积产生的网格效应，保证特征图的空间连续性，本文采用层级特征融合策略(Hierarchical Feature Fusion, HFF)对四路分支输出进行逐级累加，而非直接拼接。该策略使每一级的

输出均在前一级结果的基础上叠加当前分支特征，从而使较大膨胀率分支的输出中隐含更小感受野的上下文信息，层级累加过程表示为公式(5)：

$$K_i = \begin{cases} E_1 & i = 1 \\ K_{i-1} + E_i & i = 2, 3, 4 \end{cases} \quad (5)$$

其中 K_i 表示第 i 级累加后的融合特征，四路累加输出 K_1, K_2, K_3, K_4 在通道维度上进行拼接，经 1×1 卷积与批归一化完成特征融合与通道数对齐，最终通过残差连接与输入特征图 F 相加，保留原始特征信息并缓解梯度消失问题，该过程表示为公式(6)：

$$F_{out} = \text{BN}(\text{Conv}_{1 \times 1}(\text{Concat}[K_1, K_2, K_3, K_4])) + F \quad (6)$$

其中 $\text{Concat}[\cdot]$ 表示沿通道维度的拼接操作，残差项 F 通过跳跃连接与融合特征直接相加，确保输入输出通道数一致。

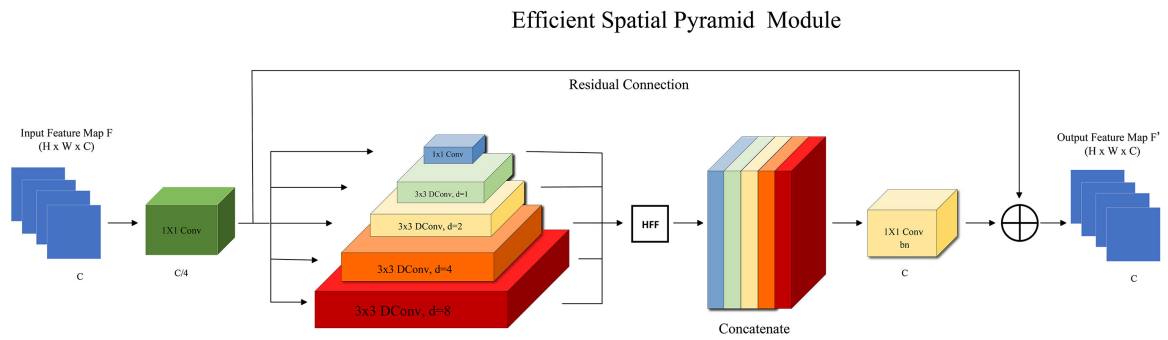


Figure 3. Architecture of the ESP module
图 3. ESP 模块结构图

2.1.3. CA 模块

CA 模块[17]基于压缩-激励网络的通道注意力思想，通过全局平均池化与全局最大池化分别生成通道统计描述符，经参数共享的多层感知器提取通道间非线性依赖关系，利用 Sigmoid 函数生成归一化的通道注意力权重，对特征图进行逐通道重标定，以增强判别性通道响应、抑制冗余特征干扰，CA 模块的结构如图 4 所示。CA 模块的具体计算流程如下：对输入特征图 $F \in R^{H \times W \times C}$ ，分别沿空间维度进行全局平均池化与全局最大池化操作，将空间信息压缩为通道描述符如公式(7)和公式(8)所示：

$$F_c^{avg} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F(i, j) \in R^{1 \times 1 \times C} \quad (7)$$

$$F_c^{max} = \max_{1 \leq i \leq H, 1 \leq j \leq W} F(i, j) \in R^{1 \times 1 \times C} \quad (8)$$

两路压缩特征分别输入参数共享的多层感知器(Multi-Layer Perceptron, MLP)，MLP 由两个全连接层构成，中间以 ReLU 激活，降维系数为 r ，用于在压缩通道维度的同时提取通道间的非线性依赖关系，通道注意力权重生成过程如公式(9)所示：

$$M_{CA}(F) = \sigma(W_1(W_0(F_c^{avg}))) + W_1(W_0(F_c^{max})) \quad (9)$$

其中 $W_0 \in R^{C/r \times C}$ 与 $W_1 \in R^{C \times C/r}$ 为两层全连接的共享权重矩阵， r 为降维系数(本文取 $r = 16$)， $\sigma(\cdot)$ 为 Sigmoid 激活函数，将输出归一化至(0, 1)区间作为各通道的注意力权重。

最终将通道注意力权重与输入特征图逐通道相乘，完成特征重标定，如公式(10)所示：

$$F' = M_{CA}(F) \otimes F \quad (10)$$

其中 $\otimes$ 表示逐通道的广播乘法操作, $F' \in R^{H \times W \times C}$ 为重标定后的输出特征图。

CA 模块引入的额外参数量仅为共享 MLP 的权重, 如公式(11)所示:

$$\Delta P_{CA} = 2 \times \frac{C}{r} \times C = \frac{2C^2}{r} \quad (11)$$

以 $C = 96, r = 16$ 为例, 额外参数量仅约为 1152 个, 相较于整体网络参数量可忽略不计, 计算开销极低。

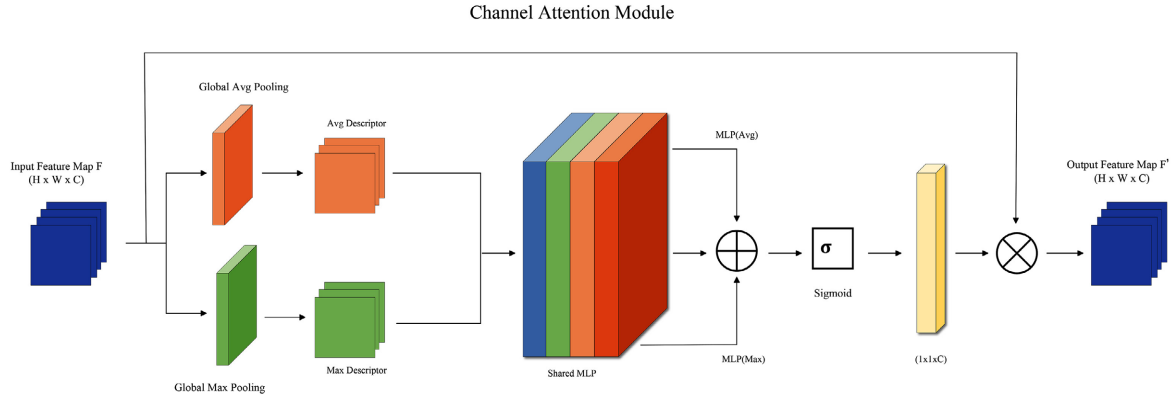


Figure 4. Architecture of the CA module

图 4. CA 模块结构图

2.2. 基于 ESP-CA-MobileViT 网络的优化策略

网络结构决定了模型的特征提取能力, 但训练策略同样对模型的收敛行为与泛化性能有重要影响。为提高 ESP-CA-MobileViT 网络的学习能力, 本文采用学习率线性衰减策略[18]、标签平滑策略[19]以及基于医学图像的迁移学习策略[20]三种优化策略的组合对网络进行训练优化, 并选用带动量的随机梯度下降算法(SGDM)作为优化器。学习率线性衰减策略使模型在训练后期以更小步长细化参数, 有效提升收敛稳定性, 避免后期震荡。标签平滑策略将硬标签软化为软标签, 在保留主要监督信号的同时为模型引入适度不确定性约束, 缓解过拟合。迁移学习方面, 采用在 ImageNet 上预训练的 MobileViT 权重初始化骨干网络, 使网络在训练初始阶段即具备基础视觉特征提取能力, 有效改善小样本条件下的泛化性能。

3. 实验设计

3.1. 实验环境及参数设置

本文实验基于 PyTorch 深度学习框架, 以 Python 为编程语言, 开发环境为 PyCharm, 硬件配置为 2 × NVIDIA Tesla P100 GPU (每卡显存 16 GB)。优化器选用 SGDM, 动量系数 0.9, 初始学习率设为 0.01, 于第 20 训练轮次线性衰减至 0.001, 总训练轮次 50, 批次大小 16, 标签平滑系数 0.01。骨干网络以 ImageNet 预训练的 MobileViT 权重初始化, 其余模块权重采用 Kaiming 正态分布随机初始化, 随后在 BraTS2019 数据集上进行全参数微调。为降低结果随机性, 所有实验均采用五折交叉验证策略, 以 5 次结果的均值作为最终性能评估指标。

3.2. BraTS2019 数据库

本文实验采用多模态脑肿瘤分割挑战赛(Brain Tumor Segmentation Challenge, BraTS)所公开的 BraTS2019 数据集。该数据集由 2019 年国际医学图像计算与计算机辅助介入会议(MICCAI)组织构建, 是脑胶质瘤分级研究中最具代表性的基准数据集之一, 已被广泛用于深度学习方法的评测与对比。

BraTS2019 数据集共包含 259 例高级别胶质瘤(HGG)和 76 例低级别胶质瘤(LGG)的多模态 MRI 扫描数据, 每例病例均提供 T1、T1C(T1 对比增强)、T2 及 FLAIR 四种模态的 MRI 序列, 每个序列包含 155 张轴向切片, 图像分辨率为 240×240 像素。图 5、图 6 为 4 种模态下的 HGG 和 LGG 肿瘤图像。四种模态在病灶信息表现上各有侧重: T1C 模态在病灶区域注射造影剂后, HGG 的增强区域信号显著增强, 肿瘤边界更为清晰, 有助于区分良恶性病灶; T2 与 FLAIR 模态则对瘤周水肿区域更为敏感。综合考量各模态的判别性信息, 本文选用 T1C 模态图像作为模型输入。

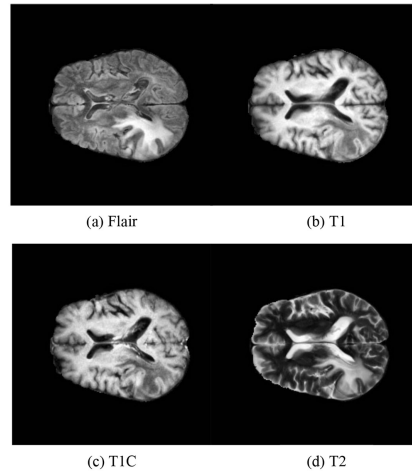


Figure 5. HGG tumor images across four modalities
图 5. 四种模态下的 HGG 图像

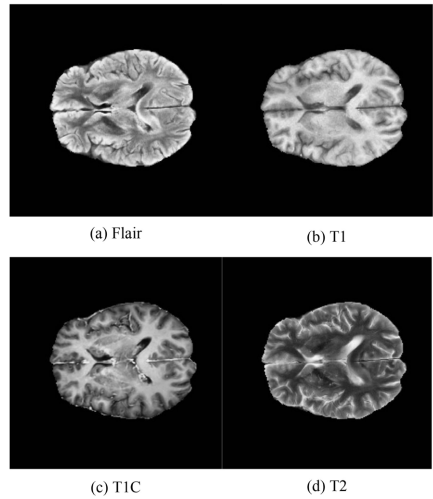


Figure 6. LGG tumor images across four modalities
图 6. 四种模态下的 LGG 图像

3.3. 数据预处理

由于 BraTS2019 数据集中 HGG 与 LGG 样本数量比例约为 3.4:1, 类别分布高度不均衡, 若直接训练极易导致模型偏向多数类。为此, 对 LGG 样本实施数据扩增操作, 包括随机水平翻转、随机旋转及随机擦除, 将 LGG 样本数量扩充至与 HGG 相当的水平。同时, 对训练集中全部样本施加随机旋转与随机擦除增强操作, 增大同一病例不同切片间的视觉差异性, 提升模型对形态变化的泛化能力。所有数据增强

操作仅应用于训练集，测试集仅执行尺寸缩放与归一化处理，以保证评价结果的客观性。

在切片提取阶段，考虑到 MRI 体积两端切片通常包含较少的肿瘤组织信息，仅保留每个病例轴向切片总数中间 60% 区间内的切片参与训练与评估。为进一步降低相邻切片间的高度相关性、减少冗余样本，采用步长为 3 的等间隔采样策略，即每隔 3 张切片抽取 1 张，在保证样本多样性的同时有效控制数据规模。为避免数据泄漏问题，本文采用病例级五折交叉验证策略，即同一患者的所有切片仅出现在训练集或测试集的一个集合内，不参与交叉混合。所有图像经预处理后统一缩放至 224×224 像素，以适配 MobileViT 骨干网络的输入规格。

3.4. 评价指标

本文从准确率(Accuracy, ACC)、灵敏度(Sensitivity, SEN)、特异度(Specificity, SPE)、阳性预测值(Positive Predictive Value, PPV)及阴性预测值(Negative Predictive Value, NPV)五个维度对分类结果进行全面评价。上述指标基于混淆矩阵计算：设 TP 为被正确判定为 HGG 的样本数， TN 为被正确判定为 LGG 的样本数， FP 为 LGG 被误判为 HGG 的样本数， FN 为 HGG 被误判为 LGG 的样本数，各指标定义如式(12)至式(16)所示。

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

$$SEN = \frac{TP}{TP + FN} \quad (13)$$

$$SPE = \frac{TN}{TN + FP} \quad (14)$$

$$PPV = \frac{TP}{TP + FP} \quad (15)$$

$$NPV = \frac{TN}{TN + FN} \quad (16)$$

3.5. 实验结果与分析

3.5.1. 消融实验

为验证 ESP 模块、CA 模块及优化策略各自的有效性，以 MobileViT 为基线，分别构造 ESP-MobileViT (仅含 ESP 模块)、CA-MobileViT (仅含 CA 模块)及 MobileViT + Opt (仅含优化策略)进行对照，并进一步设置不含优化策略的 ESP-CA-MobileViT 作为结构完整但未经训练优化的中间变体，所有变体在 BraTS2019 数据集上采用相同配置进行五折交叉验证，结果见表 1。

Table 1. Ablation study
表 1. 消融实验

模型	ACC%	SEN%	SPE%	PPV%	NPV%
MobileViT	89.09	83.17	95.00	94.56	83.45
ESP-MobileViT	94.66	91.87	97.44	97.30	92.32
CA-MobileViT	93.85	92.16	95.55	95.42	92.44
MobileViT + Opt	92.08	90.47	93.70	93.60	90.86
ESP-CA-MobileViT	96.01	93.80	98.21	98.14	94.07
ESP-CA-MobileViT + Opt	98.87	98.21	99.53	99.52	98.23

由表 1 可知，单独引入优化策略后基线 ACC 提升至 92.08%，验证了复合训练策略对小样本收敛质量的改善效果；单独引入 ESP 模块后 SEN 显著提升，说明多尺度空洞卷积有助于捕获肿瘤边界特征；单独引入 CA 模块后 ACC 与 SEN 均有提升，表明通道注意力能有效抑制冗余特征干扰。在完整结构基础上叠加优化策略后，模型各项指标均达最优，ACC 为 98.87%，验证了结构改进与训练优化策略的协同有效性。

3.5.2. 本文整体性能分析

图 7 为本文所提模型在 BraTS2019 数据集五折交叉验证中的汇总混淆矩阵。由图 7 可知，模型共正确识别 LGG 样本 7991 例，误判为 HGG 的仅 38 例；正确识别 HGG 样本 7885 例，漏诊 144 例。LGG 的识别准确率高于 HGG，这与两类样本在 T1C 模态下的影像特征差异有关——LGG 缺乏明显增强区域、特征相对单一，而部分 HGG 病灶边界模糊、与正常组织存在信号交叠，判别难度更大，导致 HGG 方向的漏诊率略高。总体而言，模型在两类样本上的误分类数量均处于较低水平，与表 2 中 SEN 和 SPE 的定量结果相互印证，表明所提方法具备较可靠的良恶性判别能力。

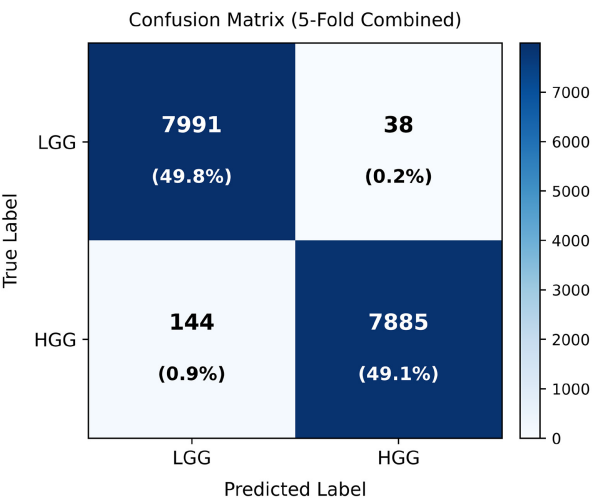


Figure 7. Confusion matrix
图 7. 混淆矩阵

3.5.3. 与其他模型的对比实验

Table 2. Comparison of classification results with other methods
表 2. 与其他模型分类结果对比

模型	ACC%	SEN%	SPE%	PPV%	NPV%	Params
MobileNetV2	86.88	84.33	95.57	98.48	64.16	12,233,920
Resnet101	87.32	86.65	89.61	96.60	66.32	68,204,610
Resnet50d	92.95	92.73	93.70	98.04	79.08	49,231,714
Efficientnetb3	93.86	93.07	96.54	98.92	80.35	25,070,378
MobileViT	89.09	83.17	95.00	94.56	83.45	7,399,714
Rubio V S <i>et al.</i> [21]	97.00	95.71	98.15	98.53	94.64	-
本文	98.87	98.21	99.53	99.52	98.23	7,614,242

为进一步评估所提方法的整体性能,将本文所提模型与 MobileNetV2、ResNet101、Resnet50d、EfficientNetB3 等经典网络及 Rubio V S *et al.*所提方法在相同数据集上进行对比,所有方法均采用相同的五折交叉验证协议与优化策略组合,结果见表 2。由表 2 可见,本文方法在 ACC、SEN、SPE、PPV、NPV 五项指标上均达到最优,分别为 98.87%、98.21%、99.53%、99.52%和 98.23%。与性能次优的 Rubio V S *et al.*相比,本文在各项指标上均有不同程度提升,且参数量仅为 7,614,242,远低于 ResNet101 等重型网络。结果表明,本文方法在保持轻量化模型结构的同时实现了更优的分类性能,具备较强的实用部署潜力。

4. 结论

本文提出一种基于改进 MobileViT 的脑胶质瘤良恶性分类方法。在 MobileViT 骨干网络基础上,引入 ESP 模块扩展多尺度感受野,融合 CA 模块对判别性通道特征进行自适应增强,并结合学习率线性衰减、标签平滑及基于医学图像的迁移学习三种优化策略,提升模型在小样本条件下的学习能力与泛化性能。在 BraTS2019 数据集上的五折交叉验证结果表明,本文方法相较 MobileViT 基线及 ResNet、EfficientNet 等经典网络在分类准确率和鲁棒性上均有所提升,ACC 达 98.87%,模型参数量仅 7.6 M,具备较好的轻量化特性。在后续研究中,可考虑引入生成对抗网络扩充医学图像训练样本,缓解小样本问题;同时探索 MRI 多序列模态信息(如 T1、T2、FLAIR 序列)的融合策略,以提供更全面的肿瘤影像特征,进一步提升模型的分类性能与泛化能力。

参考文献

- [1] Feng, J. and Yang, J. (2025) Glioma-Neuron Interactions: Insights from Neural Plasticity. *Frontiers in Oncology*, **15**, Article ID: 1661897. <https://doi.org/10.3389/fonc.2025.1661897>
- [2] Bouguerra, O., Attallah, B. and Brik, Y. (2024) MRI-Based Brain Tumor Ensemble Classification Using Two Stage Score Level Fusion and CNN Models. *Egyptian Informatics Journal*, **28**, Article 100565. <https://doi.org/10.1016/j.eij.2024.100565>
- [3] Wu, P., Wang, Z., Zheng, B., Li, H., Alsaadi, F.E. and Zeng, N. (2022) AGGN: Attention-Based Glioma Grading Network with Multi-Scale Feature Extraction and Multi-Modal Information Fusion. *Computers in Biology and Medicine*, **152**, Article 106457.
- [4] Dutta, A.K., Bokhari, Y., Alghayadh, F., Alsubai, S., Alhalabi, H.R.S., Umer, M., *et al.* (2024) Sado-Net: A Spatial Adaptive Dart Optimized Network Model for an Automated Brain Tumor Diagnosis Using MRIs. *Alexandria Engineering Journal*, **109**, 884-902. <https://doi.org/10.1016/j.aej.2024.09.102>
- [5] Falk Delgado, A. (2025) Advances of MR Imaging in Glioma: What the Neurosurgeon Needs to Know. *Acta Neurochirurgica*, **167**, Article No. 174. <https://doi.org/10.1007/s00701-025-06593-6>
- [6] Aloraini, M., Khan, A., Aladhadh, S., Habib, S., Alsharekh, M.F. and Islam, M. (2023) Combining the Transformer and Convolution for Effective Brain Tumor Classification Using MRI Images. *Applied Sciences*, **13**, Article 3680. <https://doi.org/10.3390/app13063680>
- [7] Rahman, T., Islam, M.S. and Uddin, J. (2024) MRI-Based Brain Tumor Classification Using a Dilated Parallel Deep Convolutional Neural Network. *Digital*, **4**, 529-554. <https://doi.org/10.3390/digital4030027>
- [8] Su, C., Jiang, J., Zhang, S., Shi, J., Xu, K., Shen, N., *et al.* (2018) Radiomics Based on Multicontrast MRI Can Precisely Differentiate among Glioma Subtypes and Predict Tumour-Proliferative Behaviour. *European Radiology*, **29**, 1986-1996. <https://doi.org/10.1007/s00330-018-5704-8>
- [9] Naser, M.A. and Deen, M.J. (2020) Brain Tumor Segmentation and Grading of Lower-Grade Glioma Using Deep Learning in MRI Images. *Computers in Biology and Medicine*, **121**, Article 103758. <https://doi.org/10.1016/j.compbiomed.2020.103758>
- [10] Rasool, M., Noorwali, A., Ghandorh, H., Ismail, N.A. and Yafsooz, W.M.S. (2024) Brain Tumor Classification Using Deep Learning: A State-of-the-Art Review. *Engineering, Technology & Applied Science Research*, **14**, 16586-16594. <https://doi.org/10.48084/etasr.8298>
- [11] Yang, Y., Yan, L., Zhang, X., Han, Y., Nan, H., Hu, Y., *et al.* (2018) Glioma Grading on Conventional MR Images: A

- Deep Learning Study with Transfer Learning. *Frontiers in Neuroscience*, **12**, Article ID: 804. <https://doi.org/10.3389/fnins.2018.00804>
- [12] Sajjad, M., Khan, S., Muhammad, K., Wu, W., Ullah, A. and Baik, S.W. (2019) Multi-Grade Brain Tumor Classification Using Deep CNN with Extensive Data Augmentation. *Journal of Computational Science*, **30**, 174-182. <https://doi.org/10.1016/j.jocs.2018.12.003>
- [13] 赵尚义, 王远军. 基于 3D CNN 的脑胶质瘤分类算法[J]. 光学技术, 2019, 45(6): 749-755.
- [14] Mehta, S. and Rastegari, M. (2021) MobileViT: Light-Weight, General-Purpose, and Mobile-Friendly Vision Transformer. arXiv:2110.02178.
- [15] Yu, F. and Koltun, V. (2015) Multi-Scale Context Aggregation by Dilated Convolutions. arXiv:1511.07122.
- [16] Mehta, S., Rastegari, M., Caspi, A., Shapiro, L. and Hajishirzi, H. (2018) ESPNet: Efficient Spatial Pyramid of Dilated Convolutions for Semantic Segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 561-580. https://doi.org/10.1007/978-3-030-01249-6_34
- [17] Hu, J., Shen, L., Albanie, S., Sun, G. and Wu, E. (2020) Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **42**, 2011-2023. <https://doi.org/10.1109/tpami.2019.2913372>
- [18] Defazio, A., Cutkosky, A., Mehta, H., et al. (2024) Optimal Linear Decay Learning Rate Schedules and Further Refinements. arXiv:2310.07831. <https://arxiv.org/abs/2310.07831>
- [19] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. (2016) Rethinking the Inception Architecture for Computer Vision. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 2818-2826. <https://doi.org/10.1109/cvpr.2016.308>
- [20] Alam, M.Z., Roy, T., Kawsar, H.M.N. and Rimi, I. (2024) Enhancing Transfer Learning for Medical Image Classification with SMOTE: A Comparative Study. 2024 *27th International Conference on Computer and Information Technology (ICCIT)*, Cox's Bazar, 20-22 December 2024, 245-250. <https://doi.org/10.1109/iccit64611.2024.11022326>
- [21] Valbuena Rubio, S., García-Ordás, M.T., García-Olalla Olivera, O., Alaiz-Moretón, H., González-Alonso, M. and Benítez-Andrades, J.A. (2023) Survival and Grade of the Glioma Prediction Using Transfer Learning. *PeerJ Computer Science*, **9**, e1723. <https://doi.org/10.7717/peerj-cs.1723>