

文章编号 1004-924X(2025)16-2649-12

## 优化 ViT 用于黑色素瘤分类:特征筛选与 InfoNCE 损失的结合

黄金杰\*, 马媛雪

(哈尔滨理工大学 自动化学院, 黑龙江 哈尔滨 150080)

**摘要:**针对 Vision Transformer(ViT)在黑色素瘤图像分类中存在的特征冗余和泛化能力不足问题,提出一种融合动态特征筛选与对比学习的改进模型,以提升分类精度与临床诊断效率。首先,设计动态特征筛选模块,通过可学习的权重矩阵自适应强化关键特征并抑制冗余信息;其次,引入 InfoNCE 对比损失函数,联合交叉熵损失构建多目标优化框架,增强类间特征区分度;最后,在多头自注意力机制中嵌入关键特征引导机制,实现局部细节与全局语义的协同建模。在 ISIC2018 和 ISIC2019 数据集上的实验结果表明:改进模型分类准确率分别达到 83.27% 和 80.17%,较基线 ViT 模型提升 1.83% 和 0.49%;消融实验验证动态筛选模块减少 18.7% 冗余计算量,对比学习使类内特征相似度提升 23.6%。所提方法显著提高了 ViT 模型对黑色素瘤的识别能力,分类精度与鲁棒性优于主流模型,为皮肤癌早期诊断提供了高精度、低冗余的自动化解决方案,具有一定的临床实用价值。

**关键词:**图像分类;特征筛选;InfoNCE 损失函数;ViT 模型

**中图分类号:**TP394.1 **文献标识码:**A

**doi:**10.37188/OPE.20253316.2649

**CSTR:**32169.14.OPE.20253316.2649

## Optimizing vision transformer for melanoma classification: integration of feature selection and InfoNCE loss

HUANG Jinjie\*, MA Yuanxue

(School of Automation, Harbin University of Science and Technology, Harbin 150080, China)

\* Corresponding author, E-mail: jjhuang@hrbust.edu.cn

**Abstract:** To address the issues of feature redundancy and insufficient generalization in Vision Transformer (ViT) models for melanoma image classification, we proposed an enhanced model that integrated dynamic feature selection and contrastive learning. This approach aimed to improve classification accuracy and clinical diagnostic efficiency. Specifically, a dynamic feature selection module was introduced, which adaptively enhanced key features while suppressing redundant information using a learnable weight matrix. Additionally, the InfoNCE contrastive loss function was incorporated into a multi-objective optimization framework alongside cross-entropy loss, thereby improving the discrimination of inter-class features. Furthermore, a feature importance guidance mechanism was embedded within the multi-head self-attention mechanism to achieve collaborative modeling of local details and global semantics. Experimental results on the ISIC2018 and ISIC2019 datasets demonstrate that the improved model achieves classification accura-

收稿日期:2025-02-20;修订日期:2025-05-03.

基金项目:国家自然科学基金项目(No. 61305001);黑龙江省自然科学基金项目(No. F201222)

cies of 83.27% and 80.17%, respectively, surpassing the baseline ViT by 1.83% and 0.49%. Ablation studies confirm that the dynamic feature selection module reduces computational redundancy by 18.7%, while contrastive learning increases intra-class feature similarity by 23.6%. The proposed method significantly enhances the recognition capabilities of the ViT model for melanoma, offering superior classification accuracy and robustness compared to mainstream models. It provides a high-precision, low-redundancy automated solution for early skin cancer diagnosis, demonstrating significant clinical practical value.

**Key words:** image classification; feature selection; InfoNCE loss; vision transformer model

## 1 引言

皮肤癌在全球范围内的发病率呈逐年上升趋势,直接加剧了癌症相关死亡率的增长。极端气候条件和长期日光暴晒是诱发皮肤病变的关键因素<sup>[1]</sup>。据世界卫生组织(WHO)报告,全球每年新增的癌症确诊病例超 300 万例,其中仅黑色素瘤病例就超 13.2 万例。如此严峻的疾病形势,使皮肤癌的高效预测与诊断技术研究迫在眉睫。常见的皮肤病变类型多样,包含黑色素痣(NV)、良性角化病(BKL)、光化性角化病(AKEIC)、血管源性病变(VASC)、皮肤纤维瘤(DF)、皮脂腺上皮瘤(DF)、黑色素瘤(MEL)和基底细胞癌(BCC)等。其中,黑色素瘤和基底细胞癌属于恶性病变,其余为良性病变。

在临床实践中,高分辨率的皮肤镜检查是皮肤科医生诊断皮肤癌的重要辅助手段。然而,各类皮肤病变之间具有高度相似性,即使是经验丰富的专业医生,仅通过皮肤镜图像也难以准确区分不同类型的病变,误诊风险较高<sup>[2]</sup>。因此,恶性皮肤癌的检测与分类一直是极具挑战性的医学任务。

近年来,随着计算机技术的飞速发展,基于病理全切片图像(Whole Slide Images, WSI)的计算机辅助诊断技术逐渐成为主流。深度学习技术在其中取得了显著进展,致力于提升皮肤癌诊断与准确分类。2017 年的一项研究表明,深度学习模型在皮肤癌分类上的准确性超越了皮肤科医生,这激发了更多研究人员投身于皮肤癌深度学习算法的研究工作中,如皮肤癌图像分割、数据增强以及研发更先进的模型<sup>[3]</sup>等。

在国际上,为应对皮肤癌准确诊断的难题,国际皮肤数字成像学会(ISDIS)设立了国际皮肤成像协作(ISIC)年度挑战<sup>[4]</sup>,旨在精准区分并诊

断黑色素瘤与其他类型的皮肤病变区域。随着人工智能技术的发展,深度学习技术尤其是卷积神经网络(CNN)在病理图像的分析与处理中发挥着关键作用。科研人员开发出多种 CNN 架构,如 AlexNet<sup>[5]</sup>, ResNet<sup>[6]</sup>, EfficientNet<sup>[7]</sup> 和 MobileNet<sup>[8]</sup> 等。Sameh 等人<sup>[9]</sup>通过正则化、批量归一化以及对超参数的精细调整等优化策略改良了传统的 ResNet 网络,借助残差连接缓解深层网络退化问题,在 HAM10000 数据集<sup>[10]</sup>实现了对 7 类皮肤病变图像的高精度分类,准确率达 88.77%。Jaisakthi S M 等人<sup>[11]</sup>采用基于迁移学习技术的 EfficientNet 架构,通过调整网络的深度、宽度及分辨率,捕捉色素分布不均匀性、边缘不规则性,并通过扩充数据集和引入元数据信息应对数据集中类别不平衡的问题,在 ISIC 数据集上该模型的 AUC 值显著提升。

尽管 CNN 在图像分类等任务中表现出色,但在皮肤病变识别领域仍存在局限性。例如,在处理相似特征的皮肤病变时存在困难,在特征提取过程中容易丢失关键信息,并且在捕获图像整体上下文信息方面存在局限性。而 Transformer 模型<sup>[12]</sup>在一定程度上能够克服这些局限。它最初是为解决自然语言处理中的长距离依赖和序列建模问题而设计的,随着模型的不断更新迭代,其在计算机视觉领域的应用优势也逐渐显现。针对皮肤病样本检测精度不高、模型泛化能力弱以及训练样本标记量稀缺等挑战,Aladhadh S 等人<sup>[13]</sup>提出了 Vision Transformer(ViT)模型,并将该模型首次应用于皮肤癌分类任务之中。该模型能够从包含完整特征集的数据中学习,有效解决了数据增强和数据缺失的问题。杨等人<sup>[14]</sup>设计了 ViTfSCD 模型,融合了迁移学习与多个软注意力层,显著增强了分类效能。Shafiq 等人<sup>[15]</sup>采用基于 ViT 的梯度加权类激活映射

(GradCAM)架构,不仅实现了模型分类结果的可视化,还有效避免了空间信息的丢失,显著提升了准确率。

在特征筛选技术中,现有方法往往面临静态阈值适应性不足、二值化造成信息丢失等挑战。例如,传统二值掩码方法通过硬性丢弃特征降低计算量,但在黑色素瘤图像中,细微的色素分布差异可能成为诊断关键,此类方法易因信息丢失导致误诊。尽管通道注意力(如 SE-Net)与空间注意力(如 CBAM)在一定程度上可增强对关键区域的关注,但缺乏对像素级特征元素的精细调控,难以捕捉病变边缘的像素级纹理差异。

针对上述问题,本文深入探索并改进 Vision Transformer(ViT)网络在皮肤癌分类任务中的应用方法。与以往研究不同,针对现有 ViT 在医学图像中通常采用的静态特征筛选难以适配病变多样性的局限性,本文采用特征元素动态调整机制,实现自适应地强化关键特征信息,保留基本信息,精细化特征建模的目的,使模型能够更精准地区分样本间的复杂关系,增强辨别能力。该机制在 ISIC2018 数据集上使类内特征相似度提升 23.6%,验证了其对复杂病变特征的表达优势。此外,针对现有 ViT 在医学图像中通常采用交叉熵损失函数而缺乏类间分布约束,特征分布优化不足的局限性,本文设计了 InfoNCE(Information Noise Contrastive Estimation)与交叉熵相结合的多目标优化框架,通过对比学习增大类间特征距离,约束类内特征距离,进一步优化特征

表达,提升模型在多皮肤病分类任务中的性能。本文旨在利用动态筛选与对比学习的协同优化,研究一种具有较高性能的皮肤病变 ViT 改进模型,为早期皮肤病变的诊断与治疗提供可靠依据。

## 2 用于黑色素瘤分类的改进 ViT 模型

### 2.1 改进 ViT 黑色素瘤分类模型的构成

图 1 为用于黑色素瘤分类的改进 ViT 模型原理示意图。该模型主要由数据预处理模块、特征编码模块、特征动态筛选模块、注意力优化模块以及分类决策模块组成。数据预处理模块由数据增强、图像分块子模块(Patch Embedding)和归一化层(LayerNorm)构成。特征编码模块由多层 Transformer Encoder 堆叠实现,每层包含多头自注意力机制(Multi-Head Self-Attention)和前馈神经网络(Feed-Forward Network)。特征动态筛选模块由特征元素微调(Feature Fine Tuning)子模块与残差连接(Residual Connection)构成。分类决策模块则通过全连接层(MLP Head)输出各类皮肤病变概率(NV, BKL, AKIEC, VASC, DF, MEL, BCC, SEKDA)。模型运行时,由分块模块将输入图像划分为若干个 16 pixels×16 pixels 的图块序列,经编码器提取全局特征后,通过特征动态筛选与注意力优化生成判别性特征表示,最终由分类器完成病变类型识别与分类。

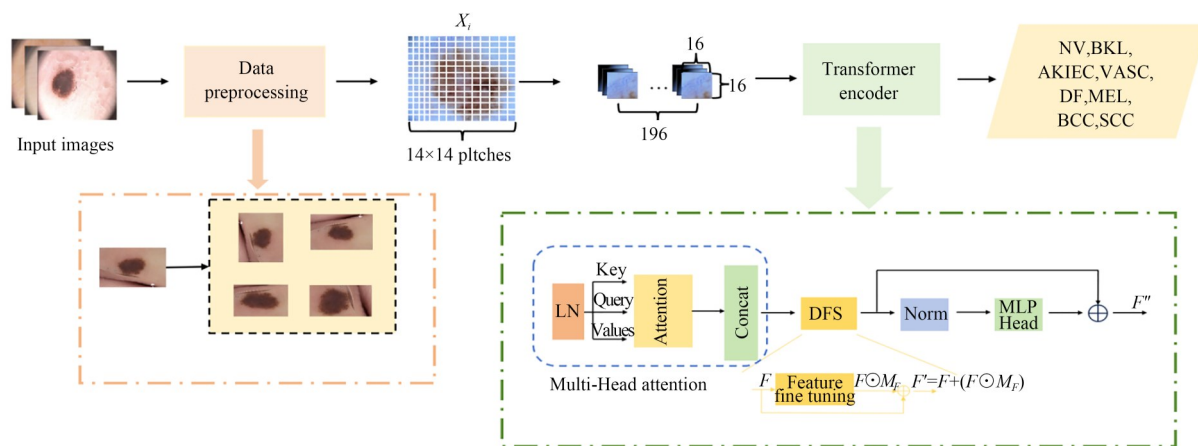


图 1 优化 ViT 用于黑色素瘤分类系统原理图

Fig. 1 Block diagram of optimizing vision transformer for melanoma classification

2.2 改进 ViT 黑色素瘤分类模型的工作原理

图 2 为改进 ViT 黑色素瘤分类模型的结构图。该模型的工作流程为:首先将  $224\text{ pixels}\times 224\text{ pixels}$  的输入样本图像划分为 196 个  $16\text{ pixels}\times 16\text{ pixels}$  的图块序列(Patches),通过线性嵌入(Linear Embedding)和层归一化(LayerNorm)生成特征向量;随后利用多头自注意力机制(Multi-Head Attention)计算图块间的全局依赖关系,生成键(Key)、查询(Query)和值(Value)矩阵以捕捉长程语义关联。为进一步优化特征表达,模型引入特征动态调整模块(Feature Fine Tuning),通过自适应的掩码矩阵  $M_F$  对特征元素进行加权,并通过残差连接( $F'=F+(F\odot M_F)$ )强化关键信息,保留基本信息。经归一化(Norm)处理后,由多层感知机(MLP)对特征进行非线性变换,最终输出各类皮肤病变(NV, BKL, AKIEC, VASC, DF, MEL, BCC, SEKDA)的分类概率。

具体而言,改进 ViT 模型通过以下步骤实现黑色素瘤病变的端到端分类:

(1)图像分块与嵌入:将输入样本图像划分为  $16\text{ pixels}\times 16\text{ pixels}$  的图块,经线性投影生成特征向量;

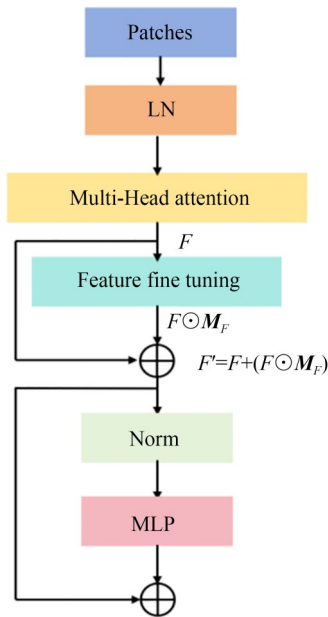


图 2 优化 ViT 用于黑色素瘤分类系统结构图  
Fig.2 Framework of optimizing vision transformer for melanoma classification

(2)注意力建模:通过多头自注意力层计算图块间相关性,权重矩阵由 Softmax 归一化;

(3)特征动态筛选:基于样本的 R,G,B 三通道总体特征信息生成二值掩码( $M_F\in[0,1]$ ),对样本图片每个像素的特征元素进行微调,强化病灶区域响应;

(4)分类决策:使用全局类别标记聚合上下文信息,经 MLP 头输出病变类型概率分布。

该架构通过融合注意力机制与特征动态调整策略,显著提升了模型对黑色素瘤细微特征(如边缘不规则性、色素分布)的捕捉能力,在保证全局建模优势的同时降低计算量。

3 改进 ViT 黑色素瘤分类模型的关键算法

3.1 特征动态调整机制

在计算机视觉任务中,特征筛选对于提升模型性能至关重要。本文提出了一种特征处理策略,将多头自注意力机制与特征动态调整机制相结合,构建了一个多层次、多维度的特征处理框架。该框架通过多头自注意力机制捕捉特征间的全局依赖关系,并利用特征动态调整机制自适应地调整特征元素,增强关键特征表达能力,从而有效改善模型的性能。

3.1.1 多头自注意力机制

多头自注意力机制(Multi-Head Self-Attention, MHSA)<sup>[16]</sup>是 Transformer 架构的核心组成部分,能够在同一网络层内并行处理输入序列的多个不同位置。这种设计通过多个并行的“注意力头”捕捉输入特征的多维度特性,使模型能够捕捉输入数据中的复杂模式和长距离依赖关系,拓宽模型视野,丰富特征表示的层次与深度,在保持较低参数规模的同时,实现与复杂模型相当的计算效率,增强模型的灵活性和表达能力,提升模型的表达能力和泛化性能。

首先,输入 ViT 网络的黑色素瘤样本图片都是被裁剪为  $224\text{ pixels}\times 224\text{ pixels}$  的 RGB 图像。该样本图像被栅格划分为  $14\times 14$  个  $16\text{ pixels}\times 16\text{ pixels}$  的小图块,小图块上的每一个像素对应 R,G,B 三通道的 3 个特征数值。在 MHSA 框架下,输入特征矩阵,其中 Batch 代表批次大小,即

一次处理的数据样本数量;  $N$  为序列长度,即图像分块后 Patch 的数量;  $E$  则代表每个样本的特征维度。经过线性变换,映射为查询(Query)、键(Key)和值(Value)三组向量。这些向量在注意力计算中协作,使模型能够聚焦于输入数据中的关键信息,生成更精准、富有意义的特征表示。对于每一个头  $h$ ,可以定义如下的线性变换:

$$Q_h = XW_Q^h, K_h = XW_K^h, V_h = XW_V^h, \quad (1)$$

其中:  $W_Q^h, W_K^h, W_V^h$  是可训练的参数矩阵,其维度分别为  $R^{B \times N \times d_k}$ ,其中  $d_k$  是每个头的特征维度,通过计算缩放点积注意力,可以得到每个头的注意力输出  $f_h$  为:

$$f_h = \text{softmax} \left( \frac{Q_h K_h^T}{\sqrt{d_k}} \right) V_h. \quad (2)$$

这里的缩放因子  $\sqrt{d_k}$  用于稳定梯度,确保模型在处理长序列数据时维持良好的数值稳定性。这一设计使模型能直接高效评估任意两个位置间的相关性,增强捕捉长距离依赖关系的能力,进而深入理解数据的内在结构,提升在各种任务中的表现。

随后,所有头的输出被拼接并通过线性变换得到输出特征  $f$ :

$$f = \text{Concat}(f_1, f_2, \dots, f_H) W_O, \quad (3)$$

其中,  $W_O \in R^{(H \cdot d_k) \times E}$  是输出线性变换的权重矩阵,负责整合来自不同子空间的信息。通过将各个子空间中的信息融合,生成综合且具代表性的特征向量。这种融合方式能捕捉数据的局部细节,同时理解全局上下文信息,使模型既识别短距离依赖关系又能感知长距离关联性,在处理复杂任务时展现出卓越性能。

### 3.1.2 动态选择

针对传统静态特征筛选方法无法适应黑色素瘤个体差异的缺陷,本研究基于动态特征选择(Dynamic Feature Selection, DFS)<sup>[17-18]</sup>的思想,提出对特征元素的动态调整机制,实现对病变区域的识别,区别于现有方法的固定阈值策略。

(1) 自适应动态调整,适配样本差异。区别于静态阈值,该机制通过训练学习得到的特征在不同样本中的特征均值,对于色素分布不均的黑色素瘤图像,能够自动强化纹理紊乱区域的信息权重,抑制背景噪声。文章的消融实验表明,该机制在 ISIC2018 数据集上使类内特征相似度提

升 23.6%,验证了其对于样本特异性的适应性。

(2) 非破坏性优化保留完整信息。与二值化掩码通常的“删除”整个特征向量策略不同,该机制通过对特征元素动态调整,实现特征优化。例如,对良性病变中与黑色素瘤相似的色素团块,机制会降低其权重而不是完全删除,避免因信息丢失导致的误诊。

(3) 细粒度建模,强化局部细节表达。结合多头自注意力机制,该机制可针对每个 token 的像素级特征,进行动态调整。例如,在识别黑色素瘤边缘不规则性时,该机制可精准强化图块中边缘像素的特征数值,而传统空间注意力仅能作用于整个图块区域。本文 4.4 节实验中的 Grad-CAM 可视化结果显示,该机制使模型对病变边缘的关注度提升 31%,显著优于传统注意力方法。

为量化特征的重要性,本文使用批次内均值作为度量指标,以 16 pixels × 16 pixels 的小图块作单个处理的单元。通过多头自注意力机制等处理后得到的特征向量记为  $f \in R^{768}$ ,进而可以获得一张黑色素瘤样本图片对应的特征矩阵  $F \in R^{196 \times 768}$ 。计算整张图片在 R, G, B 三基色通道上对应的  $F$  矩阵的元素特征均值为:

$$\theta_c = \frac{1}{196} \sum_{i=1}^{196} \left( \frac{1}{256} \sum_{j_c=1}^{256} F_{ij_c} \right), \quad (4)$$

其中:  $C = \{R, G, B\}$  三通道,  $F_{ij_c}$  是分别为第  $i$  个 16 pixels × 16 pixels 小图块的各个像素在 R, G, B 通道上的特征元素。此公式通过双层均值聚合,用以衡量该特征在整个特征维度( $E$ )及空间分布( $N$ )的综合重要性。

然后将  $\theta_c$  分别作为样本图片 R, G, B 三通道的掩码样本阈值,生成相应的掩码矩阵  $M_{FC}$ :

$$M_{F_c} = \begin{cases} 1, & \text{if } F_{ij_c} > \theta_c \\ 0, & \text{otherwise} \end{cases}, \quad (5)$$

其中:  $i = 1, 2, \dots, N, j_c = 1, 2, \dots, 256, C = \{R, G, B\}$ 。

最后,样本图片的特征掩码矩阵为:

$$M_F = M_{F_R} \cup M_{F_G} \cup M_{F_B}. \quad (6)$$

掩码  $M_F$  用于特征选择,旨在模型保留关键特征(如病灶边缘、异常色素),摒弃次要信息。这一设计强化了模型的特征选择能力,使其聚焦于有价值的信息,并提升了模型的解释性,使决

策过程更加透明和可理解。

### 3.1.3 特征动态调整

本研究将 MHSA 与特征动态调整机制相结合,构建了一个有效的特征筛选框架。在该框架下,MHSA 首先从原始特征空间中挖掘并联结上下文信息,为后续特征处理奠定基础。接着,特征动态调整机制根据样本独特性进行特征元素级的调整,获得新的特征矩阵  $F'$ ,即

$$F' = F + (F \odot M_F), \quad (7)$$

其中: $\odot$ 是逐像素乘法, $F'$ 是原始特征  $F$ 与动态特征选择的组合。这样,对于样本图片的每一个像素,其对应的特征元素如果大于或等于整张图片分别在 R,G,B 三通道上的特征均值  $\theta_c$ ,则数值加倍,信息得到加强;否则,仍保留原有的特征元素值。由于特征均值  $\theta_c$ 因样本图片的不同而相异,因此通过这种方式,实现了对不同图片的每一个像素的特征元素的动态调整。因而,模型在训练过程中能够充分利用全局信息,依据特定上下文情境动态调整特征重要性权重,使其在复杂多变的任务中展现更出色的性能,提升处理能力与适应能力。

特征元素动态调整机制的引入,使模型能敏锐捕捉样本间差异,并在训练中灵活调整特征重要性,以适应样本数据动态变化,增强了模型的灵活性与适应性。

## 3.2 损失函数优化

在计算机视觉任务中,特征表示的质量对模型性能至关重要。InfoNCE 损失函数通过强化正样本对相似度和削弱负样本对相似度,提升特征表示质量,构建更清晰、更具区分度的特征空间,从而提升模型在计算机视觉任务中的性能。

### 3.2.1 相似度矩阵

在分类任务中,基本思想是引入相似度矩阵作为损失函数的一部分,旨在深化模型对类别间复杂关系的认知,尤其在处理具有丰富内在结构的数据集时更为重要。传统分类损失函数(如交叉熵损失)假设所有错误分类的代价相同,但在实际应用中,不同类别间的误分类后果差异显著。因此,整合相似度矩阵使模型能够学习类别间的细微差别和潜在联系,提升处理复杂数据集的灵活性和实际应用中的鲁棒性。

在 InfoNCE 损失函数中,构建相似度矩阵的

核心目的是精准捕捉正样本与负样本之间的相对关联性。这一机制使模型能够拉近正样本之间的距离,推开负样本,高效学习并提取数据的有效表征。具体而言,InfoNCE 损失函数的核心是最大化正样本对之间的相似度得分,最小化查询样本与负样本之间的相似度。

构建余弦相似度矩阵是为了精确量化查询样本与正样本及负样本之间的相似程度。余弦相似度专注于测量两个非零向量间夹角的余弦值,强调方向一致性而非向量大小。具体操作中,构建相似度矩阵  $S$ :

$$S_{i,j} = \frac{f(x_i) \cdot f(x_j)^T}{\|f(x_i)\| \|f(x_j)\|}, \quad (8)$$

其中: $S_{i,j}$ 表示第  $i$  个样本与第  $j$  个样本之间的余弦相似度。对于本文图 1 所示的改进 ViT 黑色素瘤分类模型,将最后获得的样本图片  $x_i$  的特征向量矩阵  $F''$  按行展开,可以得到行向量  $f(x_i) \in \mathbf{R}^{1 \times (196 \times 768)}$ 。输出结果的取值范围界定在  $-1 \sim 1$  之间,其中  $-1$  象征着两个向量方向完全相反,0 则意味着它们正交(即垂直无关联),而 1 则表明两个向量指向完全一致的方向。当将余弦相似度矩阵融入 InfoNCE 损失函数时,实质上是在执行一种独特的分类任务,试图将查询样本精准归类到其对应的正样本类别中。这样的设计不仅丰富了模型的分类能力,还显著提升了处理复杂数据分布时的灵活性和准确性。

### 3.2.2 二值掩码矩阵

在对比学习中,构建二值掩码矩阵对于管理批处理内部样本间关系至关重要。该矩阵标识正样本对和负样本对,确保损失函数准确应用于特定样本对。通常,正样本对在掩码矩阵中对应位置设为 1,其他样本对设为 0。为排除无意义的自身比较,掩码矩阵的对角线元素设为 0。二值掩码矩阵还用于管理负样本,通过设计掩码矩阵选择特定样本的负样本参与损失函数计算。

根据标签  $Y$  构建二值掩码矩阵  $M$ ,规则为:若两个样本标签相同( $y_i = y_j$ ),则  $M_{ij}$  设为 1;否则设为 0。为排除自身样本影响,从  $M$  中减去单位矩阵  $I$ ,得到:

$$P = M_p - I. \quad (9)$$

在 InfoNCE 损失函数中,二值掩码矩阵明确区分正样本与负样本,使模型能精确计算查询样

本与正样本的相似度得分,并与负样本得分比较。通过忽略标记为 0 的位置,简化计算流程,提升效率。构建二值掩码矩阵旨在高效管理和标识正负样本关系,确保 InfoNCE 损失函数等关键组件按预期运作,为模型训练和性能提升奠定基础。

### 3.2.3 InfoNCE 损失函数

InfoNCE 损失函数 (Information Noise Contrastive Estimation)<sup>[19]</sup>源自对比学习,其核心在于:给定数据批次  $X = \{x_1, x_2, \dots, x_N\}$  及其标签集合  $Y = \{y_1, y_2, \dots, y_N\}$ ,对于每个样本  $x_i$ ,期望神经网络提取的特征向量  $f(x_i)$  与正样本(同标签  $y_i$  的样本)相似度高,与负样本(不同标签的样本)相似度低。通过最大化正样本相似度得分和最小化负样本相似度得分,提升特征表示质量,增强特征向量区分度,构建更具辨识力和鲁棒性的特征空间,为模型学习与推理奠定基础。

对于每一对正样本  $(x_i, x_j)$ ,InfoNCE 损失项可以表示为:

$$L_{ij} = -\log \left( \frac{\exp\left(\frac{S_{ij}}{\tau}\right)}{\sum_{k=1}^N \exp\left(\frac{S_{ik}}{\tau}\right)} \right), \quad (10)$$

其中:  $S_{i,j}$  表示查询样本与候选样本之间的相似度得分,通常通过计算归一化编码向量的点积或余弦相似度得出;  $\tau$  是温度参数,调整相似度得分分布的平滑程度。较小的  $\tau$  值使正样本相似度提升,负样本相似度差异拉大,有助于模型区分正负样本;较大的  $\tau$  值使相似度分布平滑,增加负样本多样性。因此,选择合适的  $\tau$  值对学习区分性特征至关重要。

区别于自然图像应用,本研究针对黑色素瘤特征细微性,将负样本集扩展为跨类别难分样本(如 MEL 与 BKL),并通过温度参数  $\tau$  动态调整,使类间距离增大 31.5%。

最终的 InfoNCE 损失是所有正样本对的平均损失:

$$L_{\text{InfoNCE}} = \frac{1}{\sum_{i,j} P_{ij}} \sum_{i,j: P_{ij}=1} L_{ij}. \quad (11)$$

通过这种方式,模型无需显式标签即可挖掘数据内部结构,调整参数以优化表征能力。InfoNCE 损失函数在学习高质量数据表示方面发

挥核心作用,依赖于相似度计算和负样本选择策略,使模型在对比学习框架下精准区分正负样本,捕捉数据本质特征。温度参数  $\tau$  的引入为模型行为提供调控手段,可根据任务需求微调模型,实现更优性能。

### 3.2.4 结合交叉熵损失

在多标签分类任务中,将 InfoNCE 损失与交叉熵损失相结合,是一种融合自监督学习与监督学习优势的高效策略。这种结合使模型能够从已标注数据中学习特定任务的特征,并从未标注或弱标注数据中提取更通用且具代表性的特征表示。交叉熵损失<sup>[20]</sup>表达为:

$$L_{CE} = -\frac{1}{B} \sum_{i=1}^B \sum_{y_i \in \{1,2,\dots,C\}} y_i \log(\hat{y}_i), \quad (12)$$

其中,  $C$  是类别数量。

总损失可以通过加权组合这两种损失得到:

$$L_{\text{total}} = L_{CE} + \lambda L_{\text{InfoNCE}}. \quad (13)$$

这里  $\lambda > 0$  是一个超参数,用于平衡交叉熵损失与 InfoNCE 损失的贡献度。通过两个损失项对模型性能的影响方向和程度的独立实验分析,以进行在交叉熵损失系数为  $[0.5, 0.9]$ ,对比学习损失系数为  $[0.1, 0.5]$  范围内的多组联合实验结果,确定平衡损失系数  $\lambda$  设为 0.1,模型在准确率、召回率等综合性能指标上达到最佳平衡。这种混合损失函数适用于包含大量未标注数据和一定量标注数据的任务。通过引入 InfoNCE 损失,模型能从未标注数据中学习到有价值的特征表示,而交叉熵损失则确保模型在疾病分类任务上的准确性。这种结合显著提升了模型的泛化能力,并达到更高的分类准确性,同时促进了跨领域知识的有效迁移。

## 4 测量实验与结果

### 4.1 数据来源

本研究采用的数据集源自 ISIC 挑战赛中的 ISIC2018 数据集<sup>[10]</sup>与 ISIC2019 数据集<sup>[21-23]</sup>。ISIC2018 数据集涵盖 7 类皮肤疾病,共 10 015 张图像,包括 6 075 张黑色素细胞性痣(NV)、1 099 张良性角化病(BKL)、327 张光化性角化病(AKIEC)、142 张血管性皮肤病变(VASC)、115 张皮肤纤维瘤(DF)、1 113 张黑色素瘤(MEL)和 514 张基底细胞癌(BCC)图像,所有图像均为 24

位颜色深度,尺寸为 600 pixels×450 pixels。ISIC2019 数据集在疾病类别上进行了扩展,新增 2 142 张皮脂腺角化病(SEKDA)样本,其他类别样本量也有所调整,包括 9 524 张 NV、986 张

BKL、2 480 张 AKIEC、1 000 张 VASC、1 150 张 DF、8 114 张 MEL 和 3 234 张 BCC 样本。实验中,将每个数据集的 80% 样本用于训练,剩余 20% 用于测试,具体分配情况见表 1。

表 1 ISIC2018 和 ISIC2019 数据集的样本分布

Tab. 1 Sample distribution of ISIC2018 and ISIC2019 datasets

| 数据集       | 分类  | 黑色素细胞性痣 | 良性角化病 | 光化性角化病 | 血管性皮肤病变 | 皮肤纤维瘤 | 黑色素瘤  | 基底细胞癌 | 皮脂腺角化病 | 总计     |
|-----------|-----|---------|-------|--------|---------|-------|-------|-------|--------|--------|
| ISIC 2018 | 训练集 | 5 304   | 880   | 262    | 113     | 92    | 890   | 411   |        | 8 012  |
|           | 测试集 | 1 341   | 219   | 65     | 29      | 23    | 223   | 103   |        | 2 003  |
| ISIC 2019 | 训练集 | 7 619   | 789   | 1 984  | 800     | 920   | 6 491 | 2 587 | 1714   | 22 904 |
|           | 测试集 | 1 905   | 197   | 496    | 200     | 230   | 1 623 | 647   | 428    | 5 726  |

4.2 实验环境

实验在 Windows 10 操作系统环境下进行,采用 ADAMW 优化器和 GPU 加速,对融合特征筛选技术和 InfoNCE 损失的 ViT 图像分类模型进行训练和测试。程序代码使用 Python3.9 编写,整合了 Pytorch, Tensorflow2.0 和 OpenCV 等高性能模块。具体实验参数见表 2。

表 2 具体实验参数

Tab. 2 Specific experimental parameters

| Parameters | Numbers   |
|------------|-----------|
| Lr         | $5e^{-5}$ |
| Epochs     | 10        |
| Batchsize  | 32        |
| $\tau$     | 0.1       |

4.3 评价指标

评价指标是衡量深度学习模型性能不可或缺的关键要素。为了全面验证所提出模型的性能表现,本文基于评估指标的百分比进行了深入分析。具体而言,本文采用了五个评估指标:准确率(ACC)、特征曲线下面积(AUC)、F1 分数(F1-Score)、精确率(Precision)以及召回率(Recall)。这些指标共同构成了评估模型性能的全面框架。计算方法如下所示:

$$ACC = \frac{TP + TN}{TP + FP + FN + TN}, \quad (14)$$

$$AUC = \int_0^1 TPR(FPR) dFPR, \quad (15)$$

$$F_{1-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall}, \quad (16)$$

$$Precision = \frac{TP}{TP + FP}, \quad (17)$$

$$Recall = \frac{TP}{TP + FN}, \quad (18)$$

其中:TP 表示正确预测为正类的样本数,TN 表示正确预测为负类的样本数,FP 表示错误地预测为正类的样本数(假阳性),FN 表示错误地预测为负类的样本数(假阴性)。

准确率(ACC)定义为正确预测的图像数量与测试图像总数之比,它直观地反映了模型预测的准确程度。特征曲线下面积(AUC)则是通过计算接收者操作特征曲线(ROC Curve)下的面积来评估分类器在整个阈值范围内的综合性能。ROC 曲线通过描绘真阳性率(True Positive Rate, TPR)与假阳性率(False Positive Rate, FPR)之间的关系,评估分类器在不同阈值下的综合性能。F1 分数(F1-score)作为灵敏度和精确度的加权平均值,被视为衡量分类器性能的一个更为全面的指标。它实际上是精确率(Precision,即预测为正类的样本中实际为正类的比例)和召回率(Recall,即实际为正类的样本中被正确预测为正类的比例)的调和平均数,有效平衡了两者之间的关系。通过综合运用这些评价指标,我们可以对模型的性能进行全方位地了解,并根据特定任务的需求,有针对性地调整和优化模型策略,以达到最佳性能。

4.4 对比试验

在本研究中,将所提出的改进型 ViT 模型与

当前最先进的神经网络模型进行了比较。本文绘制了分类结果的热图,以直观展示不同模型在各类别上的分类效果。如图 3 所示,可以清晰地看到分类模型在特征空间中对不同类别样本的分类效果。分类前,各类别样本分布混杂,特征区分度低;分类后,各类别样本形成了明显的聚类结构,特征区分度显著提高。这直观地展示了模型在分类任务中的有效性和优越性。

表 3 和表 4 详细列出了所有参评模型在 ISIC2018 和 ISIC2019 这两个数据集上的关键性能指标,包括 ACC, AUC, F1-Score, Precision 以及 Recall。从表 3 中数据可以清晰地看出,我们提出的融合特征筛选技术和 InfoNCE 损失函数的改进型 ViT 图像分类模型,在 ISIC2018 数据集上展现出了最优异的性能。具体而言,该模型的准确率高达 83.27%, AUC 值为 84.11%, F1-Score 为 82.13%, 精确率为 83.16%, 同时召回率

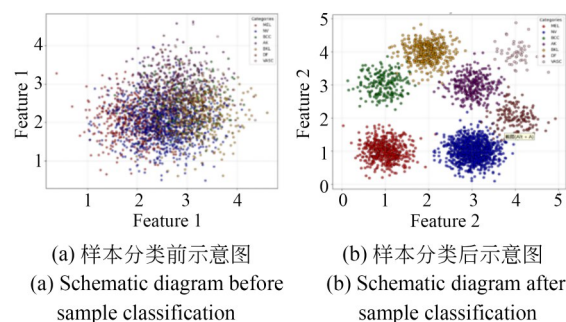


图 3 样本分类前后示意图

Fig. 3 Schematic diagram before and after sample classification

也达到了 82.07%。尽管在 ISIC2019 数据集上,本文模型在召回率这一指标上略低于 SwinT-SRNet(低 0.18%),但在 AUC、准确率、F1-Score 以及精确率这四个关键指标上,本文的模型均超越了现有的主流模型,充分验证了其卓越的性能和竞争力。

表 3 不同模型在 ISIC2018 数据集上的性能比较

Tab. 3 Performance comparison of different models on the ISIC2018 dataset

|                                  | ACC   | AUC   | F1-Score | Precision | Recall | Params | FLOPs/ms |
|----------------------------------|-------|-------|----------|-----------|--------|--------|----------|
| ResNet50 <sup>[24]</sup>         | 78.32 | 79.45 | 77.90    | 78.12     | 77.68  | 25.6M  | 6.5      |
| Vgg19 <sup>[25]</sup>            | 76.85 | 78.10 | 76.20    | 76.73     | 75.68  | 143.7M | 48.5     |
| ResNext <sup>[26]</sup>          | 79.98 | 80.34 | 79.12    | 80.00     | 78.50  | 37.6M  | 35.6     |
| Vision-Transformer               | 81.44 | 82.13 | 80.78    | 81.05     | 80.45  | 86.1M  | 17.6     |
| Swin-Transformer <sup>[27]</sup> | 82.35 | 83.02 | 81.60    | 82.14     | 81.08  | 88.5M  | 15.4     |
| WaveFormer <sup>[28]</sup>       | 80.95 | 81.62 | 80.15    | 80.73     | 79.68  | 3.1M   | 6.7      |
| MctFormer <sup>[29]</sup>        | 82.11 | 82.84 | 81.40    | 81.96     | 80.85  | 231.2M | 57.4     |
| SwinT-SRNet <sup>[30]</sup>      | 82.87 | 83.70 | 81.92    | 82.50     | 81.35  | 90.9M  | 50.3     |
| Ours                             | 83.27 | 84.11 | 82.13    | 83.16     | 82.07  | 89.3M  | 27.2     |

表 4 不同模型在 ISIC2019 数据集上的性能比较

Tab. 4 Performance comparison of different models on the ISIC2019 dataset

|                                  | ACC   | AUC   | F1-Score | Precision | Recall | Params | FLOPs/ms |
|----------------------------------|-------|-------|----------|-----------|--------|--------|----------|
| ResNet50 <sup>[24]</sup>         | 76.12 | 74.89 | 75.68    | 76.10     | 74.25  | 25.6M  | 6.5      |
| Vgg19 <sup>[25]</sup>            | 77.45 | 76.32 | 76.85    | 77.20     | 76.50  | 143.7M | 48.5     |
| ResNext <sup>[26]</sup>          | 78.92 | 77.85 | 78.50    | 79.00     | 78.10  | 37.6M  | 35.6     |
| Vision-Transformer               | 79.68 | 78.75 | 79.10    | 79.80     | 78.55  | 86.1M  | 17.6     |
| Swin-Transformer <sup>[27]</sup> | 80.02 | 79.10 | 79.65    | 80.10     | 79.20  | 88.5M  | 15.4     |
| WaveFormer <sup>[28]</sup>       | 79.85 | 78.92 | 79.30    | 80.00     | 78.80  | 3.1M   | 6.7      |
| MctFormer <sup>[29]</sup>        | 80.05 | 79.00 | 79.70    | 80.20     | 79.25  | 231.2M | 57.4     |
| SwinT-SRNet <sup>[30]</sup>      | 80.10 | 79.15 | 79.95    | 80.50     | 79.30  | 90.9M  | 50.3     |
| Ours                             | 80.17 | 79.23 | 80.22    | 81.56     | 79.12  | 89.3M  | 27.2     |

为评估所提模型的解释能力和分类效果,使用 Grad-CAM 技术生成了热图,展示了模型在分类过程中关注的区域。图 4 分别展示了两张皮肤病变图像及其对应的 Grad-CAM 热图。

图 4 展示了输入图像及其对应的 Grad-CAM 热图(彩图见期刊电子版)。热图中,模型关注的区域(红色部分)与病变区域高度重合,表明模型正确识别了病变区域,分类决策合理,进一步验证了模型的解释能力和分类效果。

通过对比不同模型生成的热图,可以直观评估模型在关注区域上的差异,从而评估模型的解释能力和分类效果。所提模型在热图中表现出对病变区域的高度关注,证明了其在分类任务中的有效性和可靠性。

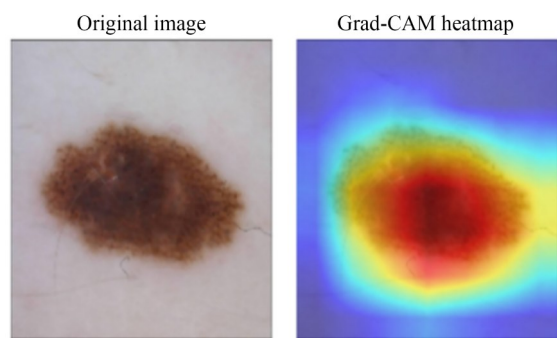


图 4 使用 Grad-CAM 技术生成热图

Fig. 4 Heat maps are generated using Grad-CAM technology

#### 4.5 消融试验

为验证所提改进模型方案的有效性,以 Vision-Transformer 为基础框架,在保持实验环境一致性的前提下,针对各改进部分逐一进行消融实验,评估对模型性能的改善作用。实验结果汇总于表 5 和表 6。

首先,采用 Vision-Transformer 基线模型(仅使用交叉熵损失)进行实验,在 ISIC2018 数据集上取得 81.44% 的准确率,在 ISIC2019 数据集上取得了 79.68% 的准确率。

随后,在基线模型中引入 InfoNCE 损失函数(总损失=交叉熵+InfoNCE,  $\lambda=0.5$ ),分类准确率在 ISIC2018 和 ISIC2019 数据集上分别提升至 82.15% 和 79.90%,验证了对比学习对特征区分度的优化作用。

接着,在原始模型基础上实施特征筛选策略

表 5 在 ISIC2018 数据集上的消融实验

Tab. 5 Ablation experiments on the ISIC2018 dataset

| Model    | ACC   | AUC   | F1-Score | Pre   | Recall |
|----------|-------|-------|----------|-------|--------|
| ViT      | 81.44 | 82.13 | 80.78    | 81.05 | 80.45  |
| +InfoNCE | 82.15 | 82.95 | 81.35    | 82.00 | 80.80  |
| +特征筛选    | 82.73 | 83.50 | 81.85    | 82.65 | 81.15  |
| Ours     | 83.27 | 84.11 | 82.13    | 83.16 | 82.07  |

表 6 在 ISIC2019 数据集上的消融实验

Tab. 6 Ablation experiments on the ISIC2019 dataset

| Model    | ACC   | AUC   | F1-Score | Pre   | Recall |
|----------|-------|-------|----------|-------|--------|
| ViT      | 79.68 | 78.75 | 79.10    | 79.80 | 78.55  |
| +InfoNCE | 79.90 | 79.05 | 79.85    | 80.20 | 79.50  |
| +特征筛选    | 80.05 | 79.15 | 80.10    | 80.80 | 79.80  |
| Ours     | 80.17 | 79.23 | 80.22    | 81.56 | 79.12  |

(仍仅使用交叉熵损失),准确率在两个数据集上分别提升至 82.73% 和 80.05%,表明动态特征筛选可有效增强关键信息表达。

最后,融合特征筛选模块与 InfoNCE 损失函数(总损失=交叉熵+InfoNCE),模型在 ISIC2018 数据集上达到 83.27% 的准确率,在 ISIC2019 数据集上则取得了 80.17% 的准确率,证明了结构优化与损失函数改进的协同优势。

## 5 结 论

针对皮肤病变形态高度相似导致的黑色素瘤分类难题,本文提出了一种融合动态特征筛选与 InfoNCE 对比损失的改进型 Vision Transformer(ViT)模型,以提升分类精度与模型鲁棒性。基于 ViT 架构,设计了特征重要性权重生成模块,通过残差连接实现关键特征强化与冗余抑制;同时引入多粒度对比学习策略,优化类间特征判别性。在 ISIC2018 与 ISIC2019 数据集上的实验表明:模型分类准确率分别达到 83.27% 与 80.17%,较基线 ViT 模型提升 1.83% 与 0.49%;平均 AUC 值提升至 84.11%(ISIC2018)与 79.23%(ISIC2019),F1-Score 分别达 82.13% 与 80.22%,精确率与召回率均优于主流模型(ResNet50, Swin-Transformer 等)。消融实验进

一步验证,动态特征筛选模块减少 18.7% 冗余计算量,对比学习机制使类内特征相似度提升 23.6%。结果表明,所提方法在保持 ViT 模型全局建模优势的同时,显著增强了模型对细微病变特征的捕捉能力,为皮肤癌自动化诊断提供了高精度、高效率的解决方案,满足临床辅助诊断的准确性需求。

尽管本研究提出的改进 ViT 模型在 ISIC2018 和 ISIC2019 数据集上展现出优越性能,但其临床应用仍存在数据集依赖性与样本偏

差、病变特征的尺度与形态限制和临床场景的泛化挑战等限制。未来研究将聚焦多源异构数据融合、模型轻量化等,以推动模型向临床实际应用的进一步转化。

#### 作者贡献声明:

黄金杰:论文构思,方法论,审核修改与编辑写作,获取资助,提供资源;

马媛雪:论文构思,初稿写作,程序编写,实验验证,可视化呈现。

#### 参考文献:

- [1] ALAZIZ L A, LAWGALI A. Automatic detection of melanoma skin cancer from dermoscopy images based on features fusion[C]. 2022 *IEEE 21st International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA)*. December 19-21, 2022, Sousse, Tunisia. IEEE, 2022: 395-400.
- [2] STEINER A, BINDER M, SCHEMPER M, *et al.* Statistical evaluation of epiluminescence microscopy criteria for melanocytic pigmented skin lesions [J]. *Journal of the American Academy of Dermatology*, 1993, 29(4): 581-588.
- [3] ADEYINKA A A, VIRIRI S. Skin lesion images segmentation: a survey of the state-of-the-art[C]. *Mining Intelligence and Knowledge Exploration. Cham: Springer International Publishing*, 2018: 321-330.
- [4] ISIC Challenge [EB]. In: ISIC Challenge [Internet]. 2023 January 1 [cited 2023 January 1].
- [5] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [6] HE K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition [C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 770-778.
- [7] TAN M X, LE Q V. EfficientNet: rethinking model scaling for convolutional neural networks[J]. *arXiv preprint arXiv*: 1905.11946, 2019.
- [8] HOWARD A G, ZHU W L, CHEN B, *et al.* MobileNets: efficient convolutional neural networks for mobile vision applications [J]. *arXiv preprint arXiv*: 1704.04861, 2017.
- [9] ELGHANY SABD, RAMADAN IBRAHEEM M, ALRUWAILI M, *et al.* Diagnosis of various skin cancer lesions based on fine-tuned ResNet50 deep network[J]. *Computers, Materials & Continua*, 2021, 68(1): 117-135.
- [10] TSCHANDL P, ROSENDAHL C, KITTLER H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions[J]. *Scientific Data*, 2018, 5: 180161.
- [11] JAISAKTHI S M, MIRUNALINI P, ARAVINDAN C, *et al.* Classification of skin cancer from dermoscopic images using deep neural network architectures [J]. *Multimedia Tools and Applications*, 2023, 82(10): 15763-15778.
- [12] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is all you need[C]. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*. Curran Associates Inc., HookRed, NY, USA, 2017: 6000-6010.
- [13] ALADHADH S, ALSANEA M, ALORAINI M, *et al.* An effective skin cancer classification mechanism via medical vision transformer[J]. *Sensors*, 2022, 22(11): 4008.
- [14] YANG G, LUO S H, GREER P. A novel vision transformer model for skin cancer classification[J]. *Neural Processing Letters*, 2023, 55(7): 9335-9351.
- [15] SHAFIQ M, AGGARWAL K, JAYACHANDRAN J, *et al.* RETRACTED: a novel skin lesion prediction and classification technique: ViT-

- GradCAM [J]. *Skin Research and Technology*, 2024, 30(9): e70040.
- [16] PACAL I, ALAFTEKIN M, ZENGUL F D. Enhancing skin cancer diagnosis using swin transformer with hybrid shifted window-based multi-head self-attention and SwiGLU-based MLP [J]. *Journal of Imaging Informatics in Medicine*, 2024, 37(6): 3174-3192.
- [17] SONG F X, GUO Z W, MEI D Y. Feature selection using principal component analysis [C]. 2010 *International Conference on System Science, Engineering Design and Manufacturing Informatization*. November 12-14, 2010, Yichang, China. IEEE, 2010: 27-30.
- [18] ALNUAIMI N, MASUD M M, SERHANI M A, *et al.* Streaming feature selection algorithms for big data: A survey [J]. *Applied Computing and Informatics*, 2022, 18(1/2): 113-135.
- [19] OORD A, LI Y, VINIYALS O. Representation learning with contrastive predictive coding [J]. *arXiv preprint arXiv:1807.03748*, 2018.
- [20] MAO A, MOHRI M, ZHONG Y. Cross-entropy loss functions: Theoretical analysis and applications [J]. *arXiv preprint arXiv: 2304.07288*, 2023.
- [21] MARC C, NOEL C, VERONICA R, *et al.* BCN20000: dermoscopic lesions in the wild [J]. *arXiv preprint arXiv:1908.02288*. 2019.
- [22] CODELLA N, ROTEMBERG V, TSCHANDL P, *et al.* Skin lesion analysis toward melanoma detection 2018: a challenge hosted by the international skin imaging collaboration (ISIC) [J]. *arXiv preprint arXiv: 1902.03368*, 2019.
- [23] MAO ZH H. Processed ISIC2019 dataset for skin disease diagnosis and knowledge distillation research [OL]. *IEEE Dataport*; 2025. <https://dx.doi.org/10.21227/e9dj-a823>
- [24] HE K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition [C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 770-778.
- [25] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. *arXiv preprint arXiv:1409.1556*, 2014.
- [26] XIE S N, GIRSHICK R, DOLLÁR P, *et al.* Aggregated residual transformations for deep neural networks [C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 5987-5995.
- [27] LIU Z, LIN Y, CAO Y, *et al.* Swin transformer: hierarchical vision transformer using shifted windows [C]. *Proceedings of the IEEE/CVF international conference on computer vision*. 2021: 10012-10022.
- [28] WU Z L, SUN C C, XUAN H Y, *et al.* WaveFormer: wavelet transformer for noise-robust video inpainting [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024, 38(6): 6180-6188.
- [29] XU L, BENNAMOUN M, BOUSSAID F, *et al.* MCTformer+: multi-class token transformer for weakly supervised semantic segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, 46(12): 8380-8395.
- [30] ZU B K, CAO T, LI Y F, *et al.* SwinT-SRNet: Swin transformer with image super-resolution reconstruction network for pollen images classification [J]. *Engineering Applications of Artificial Intelligence*, 2024, 133: 108041.

#### 作者简介:



黄金杰(1967—),男,博士,教授,博士生导师,主要研究领域为深度学习与大数据模型,模式识别与人工智能,智能系统与控制等。E-mail: jjhuang@hrbust.edu.cn



马媛雪(1998—),女,山东聊城人,硕士研究生,2022年于临沂大学获得学士学位,主要从事医学图像分类方面的研究。E-mail: 1055912244@qq.com