

# 基于深度学习的皮肤癌图像分类研究进展<sup>①</sup>

鲁 硕, 张魁星

(山东中医药大学 医学信息工程学院, 济南 250355)

通信作者: 张魁星, E-mail: [zhangkuixing@sducm.edu.cn](mailto:zhangkuixing@sducm.edu.cn)



**摘 要:** 皮肤癌是一种常见的严重癌症, 其中黑色素瘤致死率高, 而早期发现并治疗可有效提升皮肤癌患者生存率. 皮肤癌的皮肤镜图像、宏观图像、组织病理图像都是用来进行皮肤癌诊断的重要手段, 使用人工智能技术可有效提升对这 3 种图像分类的效率并降低诊断成本. 而深度学习因其具有自动提取特征的特性, 更加适应于细节丰富的皮肤癌图像分类任务. 本文综述了对皮肤癌诊断中 3 种常用图像进行分类任务的相关研究, 详细分析了 3 种图像因其图像特质不同而存在的不同的技术侧重点, 并对临床落地面临的困难进行了针对性分析. 最后, 对未来的发展与挑战进行了展望, 以进一步推动人工智能在皮肤癌诊断的广泛应用.

**关键词:** 深度学习; 皮肤癌; 图像分类; 组织病理图像; 宏观图像

引用格式: 鲁硕,张魁星.基于深度学习的皮肤癌图像分类研究进展.计算机系统应用,2026,35(2):23-39. <http://www.c-s-a.org.cn/1003-3254/10061.html>

## Research Progress on Skin Cancer Image Classification Based on Deep Learning

LU Shuo, ZHANG Kui-Xing

(School of Medical Informational Engineering, Shandong University of Traditional Chinese Medicine, Jinan 250355, China)

**Abstract:** Skin cancer is a common and serious type of cancer, with melanoma having the highest fatality rate. Early detection and treatment can significantly improve the survival rate of skin cancer patients. Dermoscopic, macroscopic, and histopathological images all play essential roles in diagnosis. The application of artificial intelligence technology can effectively enhance the efficiency of classifying these three types of images and help reduce diagnostic costs. Deep learning, with its feature extraction capabilities, is more suitable for the classification tasks of detailed skin cancer images. This study reviews the relevant research on the classification tasks of the three types of images commonly used in skin cancer diagnosis, analyzes the technical focuses of the three types of images due to their different image characteristics, and conducts targeted analysis of the difficulties faced in clinical application. Finally, future developments and challenges are discussed to promote the broader application of artificial intelligence in skin cancer diagnosis.

**Key words:** deep learning; skin cancer; image classification; histopathological image; macroscopic image

皮肤癌是一类由皮肤细胞异常增生所致的恶性肿瘤, 是全球最常见的癌症之一. 主要类型包括基底细胞癌 (BCC)、鳞状细胞癌 (SCC) 和恶性黑色素瘤 (MEL), 其中 BCC 和 SCC 通常增长较慢且局限于原发位置极少发生转移, 而 MEL 则以其高转移性和致死率而著

称. 2022 年皮肤癌新发病例为 33 万人, 死亡人数达到 5.8 万人<sup>[1]</sup>.

皮肤镜检查是一种非侵入性的影像技术, 它通过提高放大倍数、减少反射伪影来获取清楚详细的皮肤视觉数据, 有助于识别皮肤的微观结构和形态学特征,

<sup>①</sup> 基金项目: 国家自然科学基金 (62402297); 山东省自然科学基金 (ZR2023QF094); 山东省研究生优质课程建设项目 (SDYKC2023045)

收稿时间: 2025-07-11; 修改时间: 2025-08-01, 2025-08-15; 采用时间: 2025-08-29; csa 在线出版时间: 2025-11-11

CNKI 网络首发时间: 2025-11-12

广泛用于黑色素瘤和其他皮肤癌的筛查<sup>[2]</sup>. Brinker 等人<sup>[3]</sup>使用开源图像训练的卷积神经网络 (CNN) 能够在黑色素瘤皮肤镜图像分类中优于大多数皮肤科医生. 通过人工智能技术进行皮肤镜图像的自动分类可避免人工偏差, 提高诊断稳定性与效率.

相比于借助皮肤镜检查这种专业的设备, 通过肉眼辅助简单的放大与照明手段直接进行皮肤癌的诊断更为高效与便利. 通过这种方式观察到的图像数据通常借助手机或相机拍摄获得. 其图像通常带有因不均匀照明产生的过曝与伪影, 且与皮肤镜图像微观特点不同, 此类图像视角较为宏观缺失一定的细节. 但 Silveira 等人<sup>[4]</sup>的研究表明, 通过人工智能对宏观图像和皮肤镜图像进行诊断的 Kappa 值为 0.958, 表明两种方式在皮肤癌诊断方面有较好的一致性<sup>[5]</sup>, 能够借助此类图像进行初步的皮肤癌诊断. 基于人工智能技术进行宏观图像自动分类可提高皮肤癌筛查率, 提升早期发现率以降低死亡率. 但是两类图像与组织病理图像进行检验的效果差距较大, 因此组织病理图像仍为当前皮肤癌诊断的金标准<sup>[6]</sup>. 基于组织病理学检验的具体

步骤为, 组织活检后先进行切片处理, 紧接着依次采用苏木精-伊红 (H&E) 染色, 再由资深病理学专家依据国际诊断指南进行详尽、系统的形态学评估, 以确立最终的诊断结果. 但该诊断过程不仅耗时, 检查人员还需经过长时间专业的培训, 并且专家在区分良性痣与恶性黑色素瘤方面的诊断不一致性仍达 25%–26%<sup>[7]</sup>. 通过人工智能技术可避免遗漏组织病理图像信息, 并具备诊断一致性, 还可以通过可解释性手段为医生决策提供辅助与参考. 3 种图像在临床诊断中各有优势与应用面, 图 1 与表 1 展示了 3 种图像对比. 近年来, 随着深度学习技术的发展, 基于皮肤镜图像、宏观图像、组织病理图像的皮肤癌自动分类技术不断创新, 成为国内外的研究热点方向之一.



图 1 皮肤镜图像、宏观图像、组织病理图像对比

表 1 3 类图像关键指标与特点

图像	获取手段	放大倍数	分辨率	主要噪声来源	备注
皮肤镜图像	专业皮肤镜设备	10–200倍	1024×1536	毛发、血管、皮肤纹理	有行业标准及使用规范, 是医生临床诊断的主要手段
宏观图像	智能手机、相机、带有摄像头的微型设备	无标准	无标准	不均匀照明、图像模糊、衣物等非皮肤区域、毛发、肤色	无使用规范且无需借助专业设备
组织病理图像	通过病理学家采样、切片、染色等一系列操作	10倍、20倍、40倍等	分辨率极高	染色不均、健康组织占比高	皮肤癌诊断金标准. 不同厂家有其设备手册、图像质量高

本文对皮肤癌的皮肤镜图像、宏观图像、组织病理图像自动分类的相关研究进行了回顾. 对当前重点研究内容进行梳理, 对目前面临的挑战进行分析, 对未来的发展趋势进行了展望. 以期能够多角度提供深度学习技术在皮肤癌分类应用的相关内容, 为读者提供参考.

## 1 数据集

皮肤癌的皮肤镜数据集包括 ISIC 2016–2020 数据集、HAM10000 数据集、PH<sup>2</sup> 数据集, 宏观图像数据集包括 DermNet 数据集、PAD-UFES-20 数据集, 组织病理图像数据集包括 SkinCancer 数据集. 不同的数据集可以开展不同方向的研究, 以下对相关数据集进行总结.

### 1.1 ISIC 数据集

ISIC 数据集由国际皮肤影像协作组织 (ISIC) 维护, 是目前应用最广泛的皮肤病变检测数据集. 其中

ISIC 2016<sup>[8]</sup>包含 1279 张皮肤病变图像, 主要用于黑色素瘤和非黑色素瘤的分类任务. ISIC 2017<sup>[9]</sup>数据集扩展至 2750 张图像, 涵盖黑色素瘤、脂溢性角化病和良性痣 3 种病变类型. ISIC 2018<sup>[10]</sup>包含 12500 张图像, 并覆盖黑色素瘤、基底细胞癌、血管性皮肤病变等 7 类病变. ISIC 2019 和 ISIC 2020<sup>[11]</sup>继续扩展数据集, 分别包括 25331 张和 33126 张图像, 为多类别皮肤病变分类提供了更丰富的资源.

ISIC 2024<sup>[12]</sup>数据集首次完全摒弃了皮肤镜图像, 转而采用从 3D 全身摄影 (3D TBP) 中提取的标准化裁剪图像. 这些图像模拟智能手机拍摄效果, 更贴近现实中患者或初级保健医生拍摄的低质量图像, 从而提升算法在真实世界中的适应性和实用性. 该数据集包含超过 40 万张训练图像和 50 万张测试图像, 覆盖来自 3 大洲 7 家皮肤病学中心的数千名患者. 图像均为

15 mm×15 mm 的裁剪图,配有良恶性标签及患者的年龄、性别、解剖部位等元数据。相比 ISIC 2020 和更早版本主要依赖皮肤镜图像,这些图像虽然清晰、特征丰富,但通常只能在专业皮肤科诊所中获取,限制了算法在普通环境中的应用能力。此外,早期版本的数据集在病变类型和图像来源上存在一定的偏倚,往往更关注疑似恶性病变,而 ISIC 2024 则更全面地覆盖了普通良性病变,提升了模型的泛化能力。

## 1.2 HAM10000 数据集

HAM10000<sup>[13]</sup>由昆士兰大学医学院提供,包括 10 015 张皮肤镜图像,涵盖各类皮肤病变。该数据集的图像经过专家人工标注,被广泛用于皮肤癌检测研究,是 ISIC 2018 和 ISIC 2019 数据集的重要组成部分并发挥了重要作用。

## 1.3 PH<sup>2</sup> 数据集

PH<sup>2</sup> 数据集<sup>[14]</sup>由葡萄牙马托西诺斯 Pedro Hispano 医院皮肤科提供,包括 200 例皮肤镜图像,涵盖普通痣、非典型痣和黑色素瘤。PH<sup>2</sup> 数据集提供详细的医学注释和病变分割信息,常用于黑色素瘤诊断模型评估。

## 1.4 DermNet 数据集

由 DermNet<sup>[15]</sup>图像库提供,包含超过 23 000 张皮肤病变图像,是医学影像分析中最大的数据来源之一。该数据库涵盖多种皮肤病类型,并经过皮肤科专家分类,适用于人工智能医学影像研究。

## 1.5 PAD-UFES-20 数据集

PAD-UFES-20<sup>[16]</sup>由 Espirito Santo 联邦大学的研究团队构建。该数据集主要面向皮肤病诊断研究,提供了高质量的皮肤病变图像。该数据集包含了 2 298 张皮肤病临床图像,图像的分辨率为 1 050×1 050 像素。这些图像采集自不同来源,确保了数据的多样性和代表性,且每张图像均附有详细的元数据信息。数据集按照皮肤病变的类型划分为 6 个类别,其中包括基底细胞癌 845 张、黑色素瘤 52 张、痣 244 张、脂溢性角化病 192 张、日光角化病 730 张以及鳞状细胞癌 235 张。

## 1.6 SkinCancer 数据集

SkinCancer 数据集<sup>[17]</sup>是为皮肤病理图像分类的数据集,包含皮肤常见肿瘤和非肿瘤解剖组织的高分辨率图像标注。该数据集共包含 386 例病例,涵盖最常见的基底细胞癌、鳞状细胞癌和黑色素瘤,并包含多种非肿瘤组织。病理切片图像经过严格的多级人工标注,提供了 16 种不同的组织类别,包括表皮、真皮、皮下

组织、毛囊、神经、汗腺等解剖结构。该数据集中的切片图像采自德国海德堡大学病理学研究所、Trier 组织细胞与分子诊断中心以及 Hannover 皮肤病理学研究所,所有诊断均依照世界卫生组织皮肤肿瘤分类标准。

## 2 基于深度学习的皮肤癌皮肤镜图像分类研究

对于皮肤镜图像,医生通常通过 ABCD 规则、七点检查法等方法来进行诊断<sup>[18]</sup>,而深度学习可以通过自动提取相关特征完成对皮肤癌的诊断。基于深度学习的皮肤癌皮肤镜图像分类技术在提升分类准确率、可解释性、轻量化等方面取得了显著进展。

### 2.1 CNN

CNN 网络作为深度学习的经典模型,CNN<sup>[19]</sup>核心在于通过卷积核对图像进行局部扫描。这种局部感受野的设计使得网络能够捕捉到皮肤镜图像中细微的纹理、边缘和形状等信息<sup>[20]</sup>,大幅提升了皮肤癌检测的准确性与效率。

在提升分类准确率方面,皮肤癌的数据不平衡问题是因素之一。通常会通过数据增强方法进行缓解,但是这只是对原始图像机械的变换,通常会导致误差传播且仅产生有限的数据多样性。而 Zhang 等人<sup>[21]</sup>提出了类增强对比学习,设计了一个混合代理模型来生成类依赖的代理以及一种用于参数优化的周期更新策略。进一步提出平衡混合代理损失旨在利用样本和代理之间的关系平等对待不同类别样本,以防止高频类数据主导模型训练。而平衡加权交叉熵损失结合课程学习策略,提高了稀有类别样本的权重使得模型在训练阶段动态关注尾部类和诊断难度大的皮肤癌类别。对于模型优化方面,Ali 等人<sup>[22]</sup>通过两阶段的分类策略在 ISIC 2018 数据集上进行良恶性分类,准确率达到 93.4%,在两个子类进一步的分类任务中达到了 96.2% 的高精度。但是分阶段进行分类导致模型缺乏全局优化并且提高了系统的复杂度,所以应针对模型进行端到端优化。鉴于 MobileNet 网络中的传统注意力 SE 模块只关注通道特征而忽略空间信息,Cheng 等人<sup>[23]</sup>引入了 FSCA 注意力机制,该机制首先利用平均池化在水平与垂直方向上聚合特征,再结合通道注意力机制,对原始输入施加重调整,从而同时考虑了图像的空间布局和通道信息,提升模型对全局信息的捕捉能力。提出的模型对 BCC 和 NV 分类的准确率超过 80%。在超

参数调优时,传统的参数调整如果错误选择模型的超参数会导致次优结果,并且耗时颇长。Musthafa 等人<sup>[24]</sup>优化训练策略,通过 Adam 优化器进行更新模型中各层的权重参数,在 HAM10000 数据集得到了 97.78% 的准确率。Farea 等人<sup>[25]</sup>提出了一种基于人工蜂群算法的新型优化策略,通过模拟蜂群行为来寻找最优参数。并且在实验中纳入 7 个公开数据集,以改进模型在深肤色与罕见病上因训练数据不足而产生偏差的问题,该模型达到了 93% 的整体准确率。

通常通过加深网络层数以增强对图像的特征提取能力,但是当图像的细节不足时,加深网络深度并不一定能提高性能。Mehmood 等人<sup>[26]</sup>通过测试不同深度的 Xception 网络的性能,得出对于 HAM10000 数据集,更深的网络的分类效果并不是最优的。为此提出了一种基于 Xception 网络的改进的深度学习模型 SBXception,此网络降低体系结构的深度但是扩展了体系结构的广度,实现轻量化的同时进一步提高了性能。与基础模型相比,训练参数减少了 54.27%。Suleman 等人<sup>[27]</sup>使用具有可分离卷积的 MobileNet 网络以达到轻量化,同时设置不同卷积核大小及步长的多层卷积模块使得在轻量化的同时达到 89% 的准确率。Shinde 等人<sup>[28]</sup>通过提高数据集图像质量以提升检测准确率,利用黑帽滤波和 OpenCV 中的修复技术,对皮肤图像中的毛发等噪声进行去除,从而使图像更均匀,减少无关信息对模型训练的干扰。在预训练的 MobileNet 上取得了 98.9% 的准确率,并在微型嵌入式设备上进行了应用。Medhat 等人<sup>[29]</sup>提出一种基于迭代幅度剪枝的轻量化 AlexNet,该方法通过迭代方式剪除网络中影响较小的连接,以减少计算成本和内存占用。IMP AlexNet 显著减少了运行时间和内存使用,相比原始 AlexNet,平均运行时间减少约 93.33%,RAM 使用减少 40%,更加适用于手机或平板等小型设备。

对于多个表现优异的模型,通过选择其中最合适的预测结果能够降低单个模型局限性与误判风险。Hossain 等人<sup>[30]</sup>使用最大投票法将出现次数最多的结果作为最终的分类结果,但是只考虑了各个模型所输出的唯一分类结果,而没有考虑各模型输出时的置信度或概率分布。Thwin 等人<sup>[31]</sup>采用集合加权平均法将 3 种模型的预测概率结合权重进行计算,最终集合的模型准确率达到 97%,优于单个模型表现。Chatterjee 等人<sup>[32]</sup>使用网格搜索技术寻找最优的模型权重分配方案,

并在 HAM10000 数据集上达到了 98% 的准确率。Ichim 等人<sup>[33]</sup>选择多个模型通过二元模型融合方法实现多个模型融合。基于二分类子系统的策略允许每个模型专注于判断两个类别之间的差异,减少了多类别判断时可能发生的模糊和偏差,从而在整体上提升了分类的稳健性和准确性。

然而仅对结果进行决策选择忽略了模型提取的大量特征,而模型保留了从低层纹理、边缘信息到高层语义信息的丰富细节。将多个模型提取的特征进行融合,相当于构建了一个更高维、更细致的特征集,有助于形成互补,进一步提升分类效果。Ogundokun 等人<sup>[34]</sup>通过结合两个预训练的 MobileNetV2 和 Xception 模型,达到了 97.56% 的准确率。不单局限于对特征提取器进行集成的方法,Dorathi 等人<sup>[35]</sup>先通过添加 SE 注意力机制的 DenseNet 网络进行特征提取,然后使用 LSTM、ELM、SSDA 等分类器方法进行分类,最后对分类器的结果进行集成。总之,集成方法不仅充分利用了单个模型各自的优势,也实现了模型间的互补,从而在提高分类精度、增强鲁棒性等方面都展现出明显提升效果。

上述方法减少了单一模型的误差,但是集成多个模型对计算资源需求高,模型更加复杂。同时可解释性也会降低,使得难以理解每个模型的具体贡献。深度学习模型在医疗诊断场景中因黑盒特性导致信任问题,通过将可解释性技术整合到皮肤癌分类模型中,能够使医生、患者更好地理解模型决策的逻辑和依据,增加信任程度。Behara 等人<sup>[36]</sup>融合两种方法来提高局部和全局的可解释性。通过 Grad-CAM++ 方法对卷积层特征图的正偏导数进行加权组合,生成更加精确和细致的热力图,能够展示模型做出决策时关注的区域。而 FullGrad 模块通过计算整个网络中偏置项的梯度并对其求和,提供了全局视角的解释性图像。Metta 等人<sup>[37]</sup>开发了一种 ABELE 方法并通过生成范例、反范例和显著图揭示分类器的决策逻辑。ABELE 采用对抗自编码器构建潜在特征空间邻域,生成与输入图像相似或差异化的样本,并通过决策树学习局部决策规则及反事实规则以支持对分类结果的逻辑化解释,生成热力图的效果好于 Grad-CAM。

## 2.2 CNN 与机器学习融合

CNN 通过其仿生视觉系统的层次化架构,实现了从局部到全局的特征抽象。相比之下,传统机器学习算

法的核心优势在于构建高维特征空间的泛化决策边界。通过对两者优点进行集成可以进一步提高模型性能。CNN 输出的深度特征虽具有强表征能力,但其高维度与潜在冗余性可能导致分类器过拟合。机器学习模型能够在保留 CNN 特征判别力的同时优化分类边界,构建轻量级但泛化性更强的决策模型。

支持向量机 (SVM) 在处理高维数据时表现出色,尤其适用于特征数量多于样本数量的情况。它通过找到最佳超平面来最大化类间的间隔,从而提高分类精度。Abdelhafeez 等人<sup>[38]</sup>利用 ECOC 策略来训练多个 SVM 分类器处理分类问题。这些 SVM 分类器通过使用中元集技术来对不确定区域进行量化与处理。SVNSS 与 ECOC 的结合在处理不确定信息和提高多分类的准确性方面发挥了协同作用,但是对于 CNN 提取的特征并未进行处理与筛选。Olayah 等人<sup>[39]</sup>通过对 3 种 CNN 模型提取的特征进行主成分分析 (PCA) 降维后再进行融合,通过随机森林分类后在 ISIC 2019 数据集上达到 94.41% 的准确率。传统手工特征是经过医学专家经验和领域知识设计的,能够直接捕捉到医生关注的病变特征,这些特征具有更强的可解释性和临床相关性。Ahmed 等人<sup>[40]</sup>将机器学习提取的特征与深度学习模型提取的特征融合并进行 PCA 降维,输入 SVM 中完成分类,在 HAM10000 达到了 96.426% 的准确率。

上述研究仅对 CNN 提取的特征进行传统降维但并未针对目标任务进行优化。Kumar 等人<sup>[41]</sup>使用二叉树生长 BTG 算法选择不同特征尝试分类,选择表现较好的特征后继续选择与之相似的特征进行学习,表现较差的则会被剪枝优化,通过这种方式选择最佳特征组合,最后使用 SVM 完成分类。在 ISIC 2019 数据集上达到了 96.60% 的准确率。SVM 擅长找到最佳决策边界,而 K 近邻算法 (KNN) 通过邻近样本的多数投票来进行分类,适合处理多类分类问题,适用于皮肤癌多分类的场景。Zafar 等人<sup>[42]</sup>在分割图像后使用预训练的 DenseNet201 提取特征,并使用黏菌算法选择信息丰富的特征,然后将这些特征输入到 KNN 分类器中,在 ISIC 2019 数据集上获得了 91.7% 的准确率。对于寻找最优的 CNN 与机器学习组合模型,不仅追求高准确率而应多准则的综合评判。Saleh 等人<sup>[43]</sup>提出多准则决策方法评估了 51 种组合模型并使用灰狼算法进行特征筛选。基于 AlexNet 和宽型神经网络 (WNN) 的最佳模型分类准确率达到了 94.5%。通过这种量化和几何方

法,避免了主观决策中的偏差,并同时考虑了收益和成本指标。

### 2.3 Transformer

Transformer 克服 CNN 只能关注局部特征的限制<sup>[44]</sup>。在皮肤癌图像分类中,有效捕捉全局信息尤为重要,而 Transformer 的自注意力机制能关注图像中所有区域之间的依赖关系,从而聚焦与病变密切相关的区域。对比传统的卷积网络发现 Vision Transformer (ViT) 在图像特征提取和全局建模上性能更佳<sup>[45,46]</sup>。

相比 CNN 的局部归纳偏置,Transformer 的全局注意力需要更多样本学习有效的空间关系,从而实现更优的分类效果。Gallazzi 等人<sup>[47]</sup>通过整合多个小数据集构建了一个大型数据集进行训练得到了 86.37% 的准确率。不局限于更大的数据集,更小的 Patch 也能使 Transformer 的分类性能提升,Flosdorf 等人<sup>[48]</sup>采用 Patch 更小的 ViT L16,准确率达到 92.79%。与 ViT L32 相比,其性能显著提高。但是模型对黑色素瘤的召回率为 56.10%,这意味着超过 40% 的黑色瘤被归为了其他类。为了提高对黑色素瘤的敏感性,Yang 等人<sup>[49]</sup>通过叠加多个 ViT 模块提升模型对全局特征的捕捉能力,并通过层层优化增强图像特征的表达,在 HAM-10000 数据集上实现 94.1% 的准确率。得益于模型特征信息量的提高,黑色素瘤的召回率达到 68%。但是参数量远大于 ViT L16,计算资源需求高。

上述模型通过增大体积来实现良好的分类表现,然而过大的体积与参数不仅是对计算资源有高需求,还会使模型更复杂,降低模型的可解释性。Aladhadh 等人<sup>[50]</sup>使用针对医学图像优化后的 MVT 模型使得对 MEL 召回率达到了 95%。Himel 等人<sup>[51]</sup>结合最新的大模型 SAM 技术对皮肤癌图像进行分割,IoU 系数达到 96.01%,并在 HAM10000 数据集上达到 96.15% 的准确率。Remya 等人<sup>[52]</sup>使用迁移学习和通道注意力机制并融合患者元数据实现了高达 99% 的准确率和对黑色素瘤 97.8% 的召回率。同时使用通道注意力机制及可视化手段增加了可解释性。然而在真实临床数据中,通常由于数据采集困难或者设备限制,某种模态可能出现缺失,这会导致模型性能下降。Wang 等人<sup>[53]</sup>提出一种不完全模态的皮肤癌诊断系统,通过模态间和模态内的特征屏蔽来增强模型在不完全模态下的鲁棒性。该模型可以动态调整并从不同的模态组合中提取有价值的信息,而不局限于固定的模态输入。不完全模态减

轻了为每种模态组合单独设计模型的资源开销,为实际应用中的部署提供了便利。ViT可能会对背景区域过分专注,Gamage等人<sup>[54]</sup>提出了SM-ViT模型,SM-ViT结合了显著性检测器生成的显著性掩码,将其嵌入到ViT的自注意力机制中。通过这种方式,SM-ViT能够更好地聚焦于图像中最关键的前景区域,减少对背景的关注,从而显著提高了其在皮肤癌的皮肤科镜图像分类任务上的表现。

注意力机制的添加虽然能够使针对皮肤癌分类任务产生较好表现,但通常也面临模型计算资源增大的问题。Pacal等人<sup>[55]</sup>将传统的滑动窗口多头自注意力模块改进为混合式滑动视窗多头自我注意力模块,该注意力模块在传统基础上进一步扩展了窗口划分策略,不仅使用常规的正方形窗口,还引入了其他形式的分块,从而形成混合窗口机制。这种混合性令模型能够更灵活地捕捉不同方向或尺度的局部细节,更高效地处理像皮肤病灶中常见的重叠、边界模糊等问题,提取更多细粒度以及跨区域的信息,从而在不显著增加计算量的情况下,提升模型对复杂病变特征的捕捉能力。该模型在ISIC 2019数据集上的准确率为89.36%,召回率为85.13%。Yadav等人<sup>[56]</sup>通过设计了轻量级的Transformer以降低计算开销,为了不影响分类性能,设计了一个双尺度的特征提取结构以两种不同尺寸的Patch进行特征提取。既关注了局部细节,也能捕获全局上下文信息,在HAM10000数据集上达到了97.80%的准确率。Transformer的自注意力机制需要计算序列中每个元素与其他元素之间的注意力分数,计算复杂度为 $O(n^2)$ 。随着序列长度增加计算量急剧增长。Pacal等人<sup>[57]</sup>提出采用焦点自注意力机制取代传统的自注意力机制的方法。焦点自注意力结合细粒度局部交互和粗粒度全局交互的特质,不需要像传统注意力机制一样密集计算每个令牌之间的关系,从而简化了计算,最终模型仅有35.01M参数,可实现移动临床使用。Laouarem等人<sup>[58]</sup>使用一个采用傅里叶变换操作将Transformer复杂度降低到 $O(n\log n)$ 的轻量化模型FNet进行皮肤癌分类,然而因其缺乏捕获长距离依赖关系的自注意力机制导致存在较弱的特征交互和潜在的信息丢失。通过在FNet之前添加基于CNN的ASFE能够细化结构特征,确保关键病变特征得到很好的表示。FNet之后的EAM提高了特征细化和稳定性,弥补了自注意力的缺失。总的来说,Transformer轻量化提高了

临床落地的可用性,但一定程度上牺牲了准确率。

## 2.4 GAN

生成对抗网络(GAN)在2014年由Goodfellow等人<sup>[59]</sup>首次提出。该方案由两个子模型组成,一个生成器用于从域数据集生成新的可信示例,一个分类器用于将每个示例分类为真实或虚假。生成器试图通过生成看似合理的图像来欺骗用户,然后通过使用来自控制器的反馈来更新发生器的参数,从而生成真实的图像。

然而GAN结构不稳定,极易发生梯度爆炸与梯度消失。Radford等人<sup>[60]</sup>提出了一种稳定的方法,称为深度卷积生成对抗网络(DCGAN)。使用卷积层替代全连接层,这种生成器可以通过深层卷积结构学习皮肤癌图像从低级纹理到高级语义的特征,形成多尺度特征提取能力,增加了生成图像的多样性。卷积操作更适合图像数据,通过局部感受野捕捉空间相关性,跨步卷积逐步扩大生成器的感受野,从而生成了更高分辨率的图像。Behara等人<sup>[61]</sup>通过调整传统DCGAN的卷积幅度和超参数组合在ISIC 2017上达到了99.38%的准确率。DCGAN在生成对抗网络领域中起到了开创性作用,它奠定了后续GAN变体设计和发展的基础。

为了可以生成超高分辨率的图像,Saeed等人<sup>[62]</sup>首先利用ESRGAN上采样恢复细节,再通过GAN进行进一步的锐化和自然纹理合成,这种两步式方法充分考虑到医学影像中对细微结构的高要求,从而显著提升了后续分类器的泛化能力。皮肤癌数据集的不足也是导致生成图像质量不高的原因, Lee等人<sup>[63]</sup>使用PGGAN通过逐步增加网络容量,使生成器和判别器能够在低分辨率阶段先学习全局特征,再逐步加入更多细节。这样的分层特征学习机制使得即使在数据量有限的情况下,模型也能稳定收敛,生成的图像在结构、纹理以及病灶边界等方面均能高度还原真实皮肤病变的特征。Pérez等人<sup>[64]</sup>设计了一种双残差结构连接PGGAN网络,使得每个生成块接收前面所有块的输入。因此网络不仅能够捕获全局特征,还可以在深层中融合高层次信息。引入双残差结构让模型更充分地捕获复杂的病变边界和纹理特性,从而在数据增强环节生成更丰富、多样且具有鉴别性的合成图像。这对最终通过增强数据进行皮肤癌诊断具有积极意义,能够提高诊断模型对罕见或细微差异病变的辨识能力。Gowthami等人<sup>[65]</sup>添加了纹理增强模块,通过低频特征提取和深度残差稠密块来强化图像的纹理和边界信息。

采用权重归一化来替代批量归一化,提升了训练过程的稳定性并减少伪影.通过使用一个已经在面部成像训练过的 GAN 网络作为初始模型进行迁移学习,显著地减少了训练所需时间,能够生成更加真实、高质量图像.

模式崩溃是 GAN 的常见问题,主要表现为生成模型难以覆盖真实数据分布的各个样本区域.在皮肤癌数据集不平衡情况下,GAN 可能偏向生成多数类别的样本,忽视少数类别,从而导致少数类别的特定特征无法学习,生成结果趋于单一化,多样性不足,进而影响下游分类任务的性能表现. Su 等人<sup>[66]</sup>采用迁移学习的思想解决模式崩溃,先使用通用皮肤病变数据以无条件 GAN 的方式对 STGAN 进行预训练,以此学习关于皮肤图像的通用知识,再针对皮肤癌中图像较少的病变进行条件 GAN 训练.基于 STGAN 的框架在 HAM-10000 数据集上实现了 98.23% 的准确率.传统方法通过引入随机噪声来避免模式崩溃,而 Ahmad 等人<sup>[67]</sup>利用 VAE 提取了皮肤癌信息的噪声,再通过 Student T 重尾捕获一些偏离平均值的极端情况,并且还加入一个辅助分类器对生成图像进行约束. TED-GAN 框架有效拓展了生成图像的多样性和质量,分类准确率由平均 66% 上升至 92.5%.

## 2.5 CNN 与 Transformer 融合

尽管 CNN 在皮肤癌分类任务中表现优异,尤其擅长捕捉局部特征,但其局部感受野的特性使得整合全局信息方面存在局限性.这种局限性可能会妨碍对病灶形状及其与周围组织之间的关联性进行全面理解.而 Transformer 因其基于自注意力机制,能够有效建立长距离依赖关系,擅长全局特征的提取.然而 Transformer 模型计算复杂度较高,尤其在高分辨率医学图像中,可能导致计算资源消耗过大.因此结合 CNN 和 Transformer 的优点形成混合架构是一种有效的策略.通过利用 CNN 提取局部特征,同时借助 Transformer 对全局特征进行建模,可以更全面地捕捉皮肤癌病灶的上下文信息,从而提升分类性能.这种整合方式能够弥补单一模型的不足,实现性能与效率的平衡.

通过 CNN 进行特征提取之后由 Transformer 进行建模可以将二者优势得以体现. Nie 等人<sup>[68]</sup>先使用 ResNet-50 进行特征提取,提取的特征添加了位置嵌入以帮助 ViT 捕捉图像的空间分布.然后 Transformer 编码器通过自注意力机制建模长距离依赖关系,从而生成

高层次语义特征,并且验证了焦点损失函数在处理小类别样本的分类性能方面存在优势.在 ISIC 2018 数据集上达到 89.48% 的准确率,但是该模型对 MEL 病变正确分类的数量减少.为了进一步提高对 MEL 的召回率, Mateen 等人<sup>[69]</sup>使用一个组合的 Inception-ResNet 网络进行特征提取, Inception 模块可以利用多个不同尺寸的卷积核并行提取特征,从而捕捉图像中细节到全局的各种尺度信息.该混合框架在实际测试中取得了 98.65% 的准确率,对 MEL 的召回率提升到了 99.20%.

上述分阶段进行特征提取, CNN 阶段的预处理若造成信息丢失或噪声会直接影响后续 Transformer 模块的输入质量与最终效果.通过 CNN 与 Transformer 共同进行特征提取并融合,可以避免信息在前期阶段的损失与弱化. Karthik 等人<sup>[70]</sup>将 Swin Transformer 提取的全局特征与 DGSNLA 网络提取的局部特征实现融合, DGSNLA 中的 GSDW 块进行通道分组和深度卷积来提取细粒度的空间信息,同时通过通道混洗增强特征交流. ENLA 块在由 GSDW 块生成的通道增强特征图内对上下文关系进行建模,有助于捕获非局部相关性.两轨道的特征融合后通过全连接层进行最终分类,在 HAM10000 数据集上得到了 94.2% 的准确率.但是仅对良恶性肿瘤进行二分类,没有对亚型进一步研究. Ali 等人<sup>[71]</sup>利用多层感知机进行非线性融合自动选择最佳权重,显著提升了皮肤癌图像分类的准确性和鲁棒性.

虽然两种网络的集成使得模型性能得以提升,但是也使得计算资源需求增加. Abbas 等人<sup>[72]</sup>使用可分离卷积改造 SqueezeNet 网络以实现更轻量的分类模型.可分离卷积将普通的卷积分为深度卷积和逐点卷积,使得总计算量显著减少,并且在特征提取能力上与普通卷积相当. Rezaee 等人<sup>[73]</sup>减少 ResNet-18 重复堆叠的残差块并调整网络深度和宽度,使得参数量相比于 ResNet-50 约 23.5M 的体积下降至 12M. Islam 等人<sup>[74]</sup>构建了一个由 ResNet152V2、ConvNeXtBase 和 ViT Base 构建的集成教师模型,从而形成了一个更具鲁棒性和表现力的整体教师模型.使用知识蒸馏技术创建了一个仅包含 3 层卷积层、参数量极少(约 0.16M 参数、2.03 MB)的学生模型.经过 16 位量化后,最终模型甚至压缩到仅 469.77 KB,使其适合在边缘设备上高效运行.

表 2 汇总了基于深度学习的皮肤癌皮肤镜图像分类研究中各种模型性能的比较分析.

表2 基于深度学习的皮肤镜图像分类研究总结

文献	数据集	模型	分类种类数	结果	进步性	局限性
[21]	ISIC 2019	ResNet-50	八分类	准确率84.75% F1 75.13%	通过类增强对比学习提高了对于罕见病的识别能力	—
[23]	ISIC 2019	MobileNet	八分类	准确率87.3%	引入FSCA注意力机制	—
[25]	7个公开数据	Xception	九分类	准确率93.04% F1 93.12%	用ABC搜索更优初始化以缓解局部最优,提高收敛与性能	—
[26]	HAM10000	SBXception	七分类	准确率96.97%	与原始Xception模型相比参数减少了54.27%	浅层网络设计可能在其他数据集上表现不佳
[29]	PH <sup>2</sup>	AlexNet	良恶性分类	准确率96.75% 内存消耗降低40%	迭代剪枝处理	准确率相较于不进行剪枝略微下降
[30]	ISIC 2018	MobileNetV2、AlexNet、VGG16等10个CNN模型	良恶性分类	准确率93.18%	最大投票法集成结果	未考虑模型的置信度
[31]	ISIC 2018	VGG16、Inception-V3、ResNet-50	三分类	准确率97%	使用集合加权平均法赋予权重计算结果	缺乏面向结果的优化
[32]	HAM10000	InceptionV3、Xception	七分类	准确率98%	通过网格搜索寻找最优权重分配方案	—
[33]	ISIC 2019	DarkNet、GoogleNet+SVM	八分类	准确率85.74%	将复杂8类问题分解为28个二分类,利用中元集技术提升准确性	训练推理成本大,对CNN提取的特征未作处理
[40]	HAM10000	MobileNet、ResNet101+SVM	七分类	准确率96.426%	融合手工提取特征增强可解释性与准确率	—
[42]	ISIC 2019	DenseNet201+KNN	八分类	准确率91.7%	使用黏菌算法针对目标筛选特征	黏菌算法搜索代价较高,迭代次数与特征维度呈线性增长
[48]	HAM10000	SM-ViT	二分类	准确率92.79%	将显著性掩码融入ViT自注意力,增强对细粒度病灶的识别能力	—
[53]	Derm7pt	I2M2Net	七分类	相比LCKD的72.79% 提升1.23% ( $p=0.003$ )	统一模型应对任意缺模态组合	—
[58]	ISIC 2019	FNet	七分类	—	通过傅里叶变换代替自注意力机制、通过EAM保持性能	—
[64]	ISIC 2018	PGGAN	良恶性分类	准确率93% MCC 81%	引入双残差结构使生成图像质量更高	对模式崩溃未改进
[67]	HAM10000	TED-GAN	四分类	准确率92.5%	采用Student T重尾提升生成样本多样性	—
[71]	HAM10000	自定义CNN+ViT	九分类	准确率95%	通过一个多层感知机自动学习融合权重	—

### 3 基于深度学习的皮肤癌宏观图像分类研究

尽管皮肤镜图像通过显示表皮与真皮上层的微观结构提高了皮肤癌诊断的准确率,但是这种器械并没有广泛地应用在公众之中.在病情早期,通常是通过肉眼辅助普通照明设备直接观察病变部位,通过及早地肉眼观察更能够使得皮肤癌早期发现的几率增大,进而提高皮肤癌生存率.观察到的图像情况可以通过传统数码相机或者手机摄像头捕获并反映出来.通过对这种宏观图像进行人工智能自动化的分类诊断,有助于公众提高皮肤癌普查率并且可以降低对医疗资源的占用<sup>[75]</sup>.同时患者也减少了检查等待时间,隐私性也得到提升<sup>[76]</sup>.但是对于这类图像相较于皮肤镜图像,不止

是缺乏更细粒度的信息,而且通常带有因不均匀的照明产生的伪影进而导致分类困难<sup>[77]</sup>.当前的研究在宏观图像数据集构建、伪影处理和模型改进上已经取得显著进展.

由于宏观图像没有皮肤镜图像那样的高分辨率及低噪声性,所以对宏观图像的皮肤癌照片进行分类时需要获取更多的信息,以实现更高的分类准确率. Azeem 等人<sup>[78]</sup>利用多层 CNN 捕获图像数据中更为细节的纹理、颜色、轮廓信息,提出的 SkinLesNet 模型在 PAD-UFES-20 数据集上准确率达到 96%. 为了进一步提升特征提取能力, Wang 等人<sup>[79]</sup>设计了一种双分支分类框架,一个分支是基于 MobileNetV2 的原网分支,另一个

是融合网分支,原网分支负责输入图像的初始特征提取,融合网络分支使用原网提取的特征与上层输出的特征融合,并将其传递到下一层.最后对两个分支的最终结果进行加权求和.在HAM10000上达到了91.58%左右的准确率,同时在自建的采用高分辨率数码相机拍摄的皮肤癌图像数据集CSLI上也达到了79.52%准确率.但是面临模型参数大、需要更多的运算能力的问题.Chen等人<sup>[80]</sup>提出上下文特征融合模块整合来自不同深度的卷积特征来提取皮肤病图像中的细节和纹理信息,提出双重注意力模块来计算各个通道和像素的敏感性评分,再经过全连接块的调整,模型能够自动强化对关键信息的关注,并抑制背景噪声或无关区域的影响,从而提高分类的鲁棒性和准确性.在宏观图像Derm 104数据集上相比传统的基础网络最高提升了9.14%.Teymouri等人<sup>[81]</sup>将DeiT3和LightGBM两种模型进行集成,在包含来自1000多名患者的40多万张智能手机拍摄的皮肤图像数据集上对黑色素瘤识别的准确率达到89%.

对宏观图像进行研究相较于皮肤镜图像更偏向于临床初筛而非辅助医生诊断,病症与肤色多样化是初筛中面临的实际场景.皮肤癌在深色皮肤上可能不如浅色皮肤表现明显,医生在诊断时容易忽略早期迹象,因此模型在深肤色人群皮肤癌诊断时也容易产生一定的偏差.Wang等人<sup>[82]</sup>通过结合患者元数据以提升跨肤色分类能力,通过使用逐元素乘法融合模型将输入映射到高维非线性特征空间来捕获复杂的特征交互,使得无需显著增加网络宽度,就能够更好地对深肤色人群重点关注以改善分类性能.Yan等人<sup>[83]</sup>提出的PanDerm在庞大数据集上训练后能够识别多种罕见病并减少深肤色人群的诊断偏差.PanDerm使用遮掩潜变量建模和CLIP特征对齐技术进行自监督学习,能够在罕见病少量学习的情况下促进对数据结构和潜在特征的深刻理解,改善了罕见病的诊断能力.宏观图像相较于皮肤镜图像通常具有较大的背景影响模型的诊断,通常在临床诊断过程中,皮肤科医生通过逐步缩小观察范围以确定核心病灶.Zhou等人<sup>[84]</sup>提出通过模拟医生动作以有效捕捉这些视觉特征以提高分类准确率与用户信任度,使用了ViT作为共享主干.分别使用3个独立的任务特定头来预测疾病、身体部位和属性,每个头包含两个独立的ViT层,每一层的提取特征被分为图像特征和斑块特征,将两组特征输入到病灶选择模块中

以选择信息量最大的病灶进行标记.最后将疾病、身体部位和属性的特征表示传送到交叉交互模块以生成更全面的表示,用于最终鉴别诊断.

宏观图像对于罕见病与深肤色分类能力不足的根本还是数据短缺影响模型训练,Wang等人<sup>[85]</sup>利用扩散模型生成少数类别肤色数据.结合Textual Inversion技术捕获细腻的皮肤病特征,并通过LoRA机制对模型进行优化以增强视觉细节.但Gottfrois等人<sup>[86]</sup>提出传统生成模型可能在生成一张深肤色皮肤上的黑色素瘤图像时包含一些不真实的特征,可能出现真实世界不存在的可见的红斑和低色素边界.在生成一张浅肤色皮肤上的痤疮图像时,也可能生成几乎只会在深色皮肤类型中出现的剩余色素沉着.并且传统生成图像需要不断监督和调整超参数,原始训练集中深肤色人群占比较低时难以生成多样的皮肤癌图像,并且缺乏表现疾病物理和形态特征的能力.Kim等人<sup>[87]</sup>提出通过基于知识来合成宏观图像,这种技术能够分层控制生成内容融合为皮肤癌图像,生成的图像更加真实.随着合成实例的加入,深色肤色的处理性能得到改善.

为了扩大初筛范围以提高普查率,对于宏观图像分类的模型应具备更小的资源占用以适应移动设备、微控制器与嵌入式设备的限制.在具体实施环境中,还可能受限于网络资源导致无法访问云数据.为了解决这类问题,Kabir等人<sup>[88]</sup>提出了一种优化皮肤癌检测的轻量化方法,通过知识蒸馏将ResNet-50和DenseNet161的知识整合至TinyStudent模型,同时结合多教师学习、模型拼接和模型汤等技术,提升预测性能并减少计算资源需求.最终利用Django框架开发了一款网页应用程序,它可以在本地执行图像分类任务而不需要上传数据到云端.同样可以实现此类功能的还有TinyML技术,TinyML是一个机器学习领域的分支,专注于在资源受限的小型设备上运行AI模型,它具备超低功耗因此非常节能,且能够本地处理数据无需依赖云端减少了延迟并提高了隐私保护.Watt等人<sup>[89]</sup>使用在ISIC 2020数据集上训练达到85%准确率的MobileNet V2模型.经过10个轮次训练后迁移至Raspberry Pi 3 Model B进行移动化部署,最终在英国国家卫生服务网站上的皮肤病变图像进行了相关测试,模型显示了每种预测皮肤癌病变类别的概率分数.未来通过宏观图像进行皮肤癌分类的落地应将更多的轻量化模型通过

TinyML 这类技术以实现移动部署,进而探索更多落地应用场景的可能性。

对于迁移学习的使用, Carse 等人<sup>[90]</sup>构建了一个由初级保健人员采集的图像与转诊至二级护理后由经过培训的医护人员采集的图像构建的两种数据集. 并使用迁移学习以改进小数据集上性能不佳的问题. 结果发现使用皮肤镜图像进行预训练对宏观图像分类并没有带来额外优势. 这显示出数据类型一致性的重要性, 也证明对宏观图像的分类研究与皮肤镜图像存在不同, 需要进一步研究以针对宏观图像. 为了能有效发挥迁移学习的作用, Somfai 等人<sup>[91]</sup>利用大规模的多源数据集 (包含宏观图像) 进行训练后, 经过迁移到使用宏观图像训练的模型上可以取得大约 79% 的准确率. 并且发现通过将多个关注不同特征和采用不同训练策略的模型组合起来, 能够有效弥补单一模型在不同图像采集条件下的表现不足.

为了减少宏观图像产生的阴影和过度曝光, Santos 等人<sup>[92]</sup>采用多种深度学习技术对宏观图像进行光照与颜色的调整. 使用 LCDPNet 解决曝光不足与过度曝光

问题, LCDPNet 采用专门的 LCDE 模块来捕获图像中的局部颜色特征对图像进行校正. 使用 LLNeRF 技术利用神经辐射场的思想对低光图像进行增强, 通过无监督的方式分解光照信息来校正低光图像中的噪声、颜色失真以及细节复原问题. 使用 DSN 技术结合递归学习与注意力机制来逐步调整图像中的低光照区域、校正颜色失真以及恢复图像内容的细节. 其核心优势在于能够双向提取图像特征, 从而在渐进式的学习过程中避免一次性处理所有修正任务所带来的误差, 提升最终图像的视觉质量和细节保留. 应用多重技术后, 模型在精确率上平均提升了约 3.4%. 但是存在少数模型准确率下降的情况, 说明使用深度学习进行图像校正, 具有一定的不可靠性. 针对宏观图像低分辨率的特点, Al Mahmud 等人<sup>[93]</sup>提出的 SkinNet-14 通过改进紧凑卷积 Transformer (CCT) 架构, 实现了仅使用 32×32 像素的图像也能进行高效的分析, 在 PAD-UFES-20 达到了 98.14% 的准确率.

表 3 汇总了基于深度学习的皮肤癌宏观图像分类模型性能的比较分析.

表 3 基于深度学习的皮肤癌宏观图像分类研究总结

文献	数据集	模型	分类种类数	准确率	进步性	局限性
[79]	自建数据集CSLI	MobileNetV2	六分类	79.52%	双分支融合网络加强特征提取能力	模型参数大
[81]	ISIC 2024	DeiT3+LightGBM	良恶性分类	89%	集成模型发挥各模型优势	—
[83]	11个临床机构数据集	ViT	二分类	超过次佳模型 9% ( $p < 0.001$ )	通过遮掩潜在变量提升了对罕见病的学习能力	—
[85]	Fitzpatrick17k 数据集	ViT	七分类	88.7%	采用了基于扩散的生成框架扩展数据集	会导致生成不真实的图像
[88]	PAD-UFES-20	DenseNet161、ResNet-50	七分类	85.7%	知识蒸馏轻量化模型并无需联网本地部署程序	—
[90]	自建数据集	EfficientNet+ SwinTransformer	七分类	91%	使用皮肤镜图像进行迁移学习会导致效果不佳	—
[91]	多数数据集 (PAD-UFES-20、DermNet)+私人数据集	自定义CNN	二分类	79%	结合多源数据集有效减少数据集依赖, 并提升了泛化性能	—
[92]	PAD-UFES-20	ResNet-50	良恶性分类	92.9%	引入了LLNeRF等技术进行光照处理, 平均精度提升了约3.4%	图像校正有一定不可靠性
[93]	PAD-UFES-20	SkinNet-14	三分类	98.14%	针对宏观图像低分辨率进行适配	—

#### 4 基于深度学习的皮肤癌组织病理图像分类研究

全切片图像 (WSI) 通常是超高分辨率的图像, 在深度学习模型训练时, 直接输入整个 WSI 会影响数据加载和处理速度. 通常做法是先对 WSI 划分为更小的区域 (称为瓦片), 通过对瓦片进行处理来完成对整个 WSI 的分类. 同时皮肤癌组织病理图像的标注是费时

且稀缺的, 尤其是在对 WSI 划分后会生成大量的瓦片, 对它们逐个进行标注变得不现实. 采用弱监督的多实例学习 (MIL), 是指在不需要对瓦片进行标注只知道 WSI 类别的情况下, 能够通过对每个瓦片进行概率预测, 并筛选出一批瓦片进行信息聚合, 从而实现对整个 WSI 类别的预测, 显著降低标注成本并提升模型的泛化能力和可解释性, 具体流程如图 2 所示.

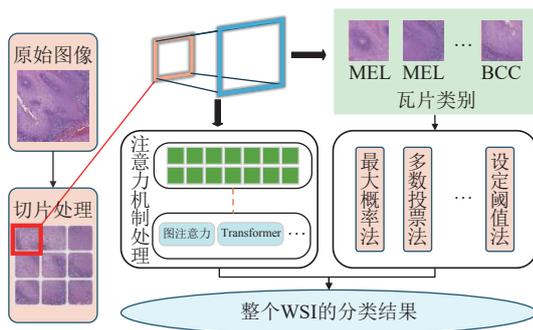


图2 多实例学习流程图

#### 4.1 基于传统多实例学习的皮肤癌组织病理图像分类

多实例学习中是通过多个瓦片预测整个 WSI 的类型, 如何更好地聚合瓦片预测的结果和提高模型泛化能力成为当前研究重点. Li 等人<sup>[94]</sup>选取了多种不同尺寸的瓦片并使用 MIL 方法训练 CNN 网络, 在对瓦片聚合时采用最大概率法. 最终发现 512×512 像素尺寸的模型表现最佳, 在 WSI 层面和瓦片块层面上都具有较高的准确性. 但是该模型仍存在一定的误诊, 模型将 Spitz 痣或蓝痣错误分类为黑色素瘤. 考虑到因为最大概率法策略导致更倾向于检测最具特征性的病灶, 因个别瓦片的错误分类导致误判. Fragoso-Garcia 等人<sup>[95]</sup>首先对病理图像进行分割瓦片出可能是肿瘤部分的区域. 随后使用 EfficientNet-B5 模型对每个瓦片进行单独产生分类结果后再使用多数投票法得出最终分类结果. 通过多数投票法更稳定, 但是当少数瓦片显示病灶但未达到多数投票标准时, 导致忽略了关键小区域致使分类错误. 不同放大倍数的组织病理图像提供的信息是不同的, Chaurasia 等人<sup>[96]</sup>使用了多分辨率图像训练一个多分类 ViT 模型, 使得提高了模型在不同分辨率上捕获特征的能力. 在瓦片信息聚合层面, 如果超过 5% 的瓦片被预测为癌症类别, 则整个 WSI 被归类为癌症. 这种方式相比较于多数投票法, 更能提高检测癌症的灵敏度, 但也导致假阳性率的增加. 由于资源设备限制会导致获取的组织病理图像质量不高, 从而导致分类效果不佳. Feng 等人<sup>[97]</sup>探究了 ViT 在低质量鳞状细胞癌组织病理图像上的应用, 添加批归一层稳定训练过程, 全连接层以强化特征学习. 验证了 ViT 的表现是超过 CNN 的, 但是也提高了对于资源的需求.

#### 4.2 基于注意力机制多实例学习的皮肤癌组织病理图像分类

上述仅对单独瓦片进行高维特征提取, 并通过传统瓦片聚合策略以预测整体分类效果. 但是在皮肤癌

实际诊断中, 病理学家不仅会关注局部的病变特征, 还会结合整个 WSI 的信息来判断病变类型. 所以对同属于一个 WSI 的瓦片进行全局建模使模型能够充分获取全局信息尤为重要. Shi 等人<sup>[98]</sup>使用两阶段方法, 第 1 阶段采用 CNN 对所有瓦片得出类别的预测概率, 在第 2 阶段选取概率较高的瓦片再通过 AGGN 建模空间关系, 使得能够把握对整体病灶信息的捕获, 从而得出更准确的分类结果. 同时这种方法仅把部分瓦片纳入第 2 阶段进行特征提取, 大大减少了对算力的需求. 然而在第 1 阶段需要对大量的组织病理瓦片进行处理以获取感兴趣的病变区域, 但组织病理图像中健康区域可能占比更大, 所以对健康区域的识别造成了大量的资源浪费. Nejat 等人<sup>[99]</sup>通过 kimiaNet 模型学习健康组织图像, 提出了一种异常检测方法从而剔除健康组织冗余信息, 使得模型更加专注于病变区域. 通过在皮肤鳞状细胞癌数据集上测试后, 发现正常组织可以减少瓦片数量 67%, 同时保持索引和搜索性能不变. Geijs 等人<sup>[100]</sup>对比了集群约束注意力多实例学习 (CLAM) 与 StreamingCLAM, CLAM 通过多头注意力机制计算每个瓦片的重要性并赋予权重. 而 StreamingCLAM 可以对整个 WSI 图像进行处理, 采用流式处理的思想, 不再先进行瓦片划分而是分为连续的小块输入网络中. 在局部卷积和池化操作的帮助下, 块在网络的较深层次进行拼接, 从而形成全局、连续的特征. 这种方式既保留了全局上下文信息, 又能在内存允许范围内动态扫描整个 WSI. Björkman 等人<sup>[101]</sup>使用预训练的 SimCLR 模型对每个瓦片提取高维特征, 随后按照瓦片的原始空间位置构建图结构, 充分利用了相邻区域的上下文信息, 更准确地反映了整个瓦片的组织结构. 构建好的图结构通过图注意力和 ViT 进行了全局信息的提取.

#### 4.3 结合其他技术在皮肤癌组织病理图像分类中的应用

在皮肤癌组织病理图像分类中, 结合一些其他技术能够使得在实际生活中的应用场景增加. Tran 等人<sup>[102]</sup>将文本信息与病理信息通过跨注意力机制进行融合, 采用了基于 BioGPT 的语言模块, 实现了从 WSI 中自动生成皮肤病理报告. 与皮肤癌类似的皮肤病变种类繁多, 模型应不断学习新的皮肤病变知识以避免误分类. 但是当模型不断学习新数据时, 容易遗忘之前学到的信息. 为了解决这一问题, Meseguer 等人<sup>[103]</sup>提出了一种嵌入匹配损失的知识蒸馏方法, 使学生模型保持教师模型的特征分布一致性, 强化对旧类别知识的保

持. 实现了将 MIL 与类增量学习相结合, 为真正的复杂皮肤癌诊断情况提供了解决方案.

表 4 总结了基于深度学习的皮肤癌组织病理图像模型分类性能的比较分析.

表 4 基于深度学习的皮肤癌组织病理图像分类研究总结

文献	数据集	模型	分类种类数	准确率 (%)	进步性	局限性
[94]	90张WSI自建数据集	EfficientNetB1	良恶性分类	82.1	最大概率法聚合瓦片结果	易因为个别瓦片错误分类而误判
[95]	450张WSI自建数据集	EfficientNet-B5	良恶性分类	95	多数投票法更稳定	忽视关键小区域, 召回率较低
[98]	自建数据集	EfficientNet B6+AGGN	四分类	95.78	利用AGGN实现了切片级空间信息聚合, 显著提高诊断性能	—
[100]	内部数据集 COBRA	ResNet-39	三分类	99.7	StreamingCLAM改进了特征提取过程, 可以直接在整个WSI上进行端到端的训练	—
[101]	内部数据集	ResNet18+ GCN、ViT	两分类	91.7	通过图注意力根据周围邻节点的信息汇总特征, 实现对图中肿瘤边界及组织模式的有效检测	—

## 5 讨论与展望

当前深度学习在皮肤癌分类领域取得了显著进展, 不断改进的模型与新技术的使用使得分类的准确率与泛化性不断提高. 但是当前仍然面临着诸多挑战与困难.

(1) 单一模型在面对数据时可能存在其固有偏差, 但当多个模型协同工作时, 它们的误差可以相互抵消从而提升整体的预测精度和稳定性, 因此集成模型在皮肤癌 3 种图像分类任务中的表现几乎都优于单一模型. 但集成模型面临可解释性降低与模型复杂度较高的问题.

针对上述问题, 未来可通过知识蒸馏对子模型轻量化处理以降低资源消耗, 并且通过 LIME 等技术对各个子模型的贡献进行量化和可视化, 从而揭示整体决策过程, 以便提高医生与患者对人工智能的信任度.

(2) 皮肤癌宏观图像分类模型如果仅在迁移学习中使用皮肤镜图像进行预训练, 会导致模型最终表现不佳. 同时, 使用高分辨率图像训练宏观图像分类模型会导致泛化性不强, 因为现实中智能手机拍摄的图像受限于使用场景通常是带伪影、高噪声、模糊的.

因此, 针对宏观图像应构建专门的预训练与训练策略. 应建立模拟智能手机真实应用场景的大型皮肤癌宏观图像数据集. 对真实场景中影响图像的因素应进行处理与分析, 以便减少对模型的干扰.

(3) 对于通过深度学习技术校正宏观图像, 可能会导致分类模型的准确率下降. 造成这种情况的因素很多. 首先, 在校正过程中, 为了消除光照不均和颜色偏差, 算法可能会过度调整图像. 这种过度校正可能会模糊或削弱原本包含判别信息的细微特征, 而这些细节对某些模型来说是区分恶性和良性皮肤肿瘤的重要依据. 其次, 在校正过程中, 如果处理不当可能会导致噪声或

失真, 噪声的引入会干扰模型的特征提取过程, 进而降低分类器对病变特征的判别力, 导致整体准确率下降.

针对上述情况, 未来的发展可以将照明与颜色校正模块嵌入整个分类网络, 形成一个端到端的优化系统, 让校正模块、特征提取和后续分类的目标函数共同优化. 这样校正模块可以学习到如何调整输入图像, 在保留关键信息的同时消除不必要的干扰, 从而提高整体准确率.

(4) 轻量化对于皮肤癌筛查应用部署在移动设备上以实现高普查率尤为重要, 知识蒸馏作为一种重要的模型压缩方法, 它将一个预先训练好的复杂大模型(教师模型)的知识传递给一个轻量级的小模型(学生模型), 使后者在参数量大大减少的情况下依然能够达到接近教师模型的性能. 但是教师模型中隐含的知识可能难以被完美提取和传递, 从而导致学生模型未能完全恢复教师模型的潜在能力.

为解决知识蒸馏中教师模型隐含知识难以传递的问题, 可以采用多层次特征蒸馏策略, 结合中间层特征、注意力图等信息增强学生模型的表达能力, 同时可使用多教师协同蒸馏和对比学习等方法进一步丰富知识来源以提升轻量模型在低资源环境下的准确性和泛化能力.

(5) 当前人工智能在皮肤病理图像诊断中仍然受限于数据集的训练范围与多样性. 现有的人工智能模型通常基于特定病症的分类任务进行训练, 缺乏对复杂、多病灶或非典型皮肤病变的充分识别能力. 此外, 由于患者可能同时存在多种皮肤病变, 或某些疾病表现出不典型的病理特征, 这进一步加剧了诊断的挑战.

为解决人工智能在皮肤病理图像诊断中受限于数据集范围与多样性的问题, 可以通过构建更大规模、

跨机构、多种族和多病种的高质量标注数据集来增强模型的泛化能力. 同时引入多任务学习框架, 使模型能够同时识别多种皮肤病变, 提升对复杂和非典型病灶的识别能力. 并且可以引入临床元信息 (如患者年龄、病史等) 进行多模态融合, 也有助于提升模型对真实临床场景的适应性和诊断准确性.

### 参考文献

- 1 Bray F, Laversanne M, Sung H, *et al.* Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 2024, 74(3): 229–263. [doi: [10.3322/caac.21834](https://doi.org/10.3322/caac.21834)]
- 2 Haenssle HA, Fink C, Schneiderbauer R, *et al.* Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of Oncology*, 2018, 29(8): 1836–1842. [doi: [10.1093/annonc/mdy166](https://doi.org/10.1093/annonc/mdy166)]
- 3 Brinker TJ, Hekler A, Enk AH, *et al.* Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *European Journal of Cancer*, 2019, 113: 47–54. [doi: [10.1016/j.ejca.2019.04.001](https://doi.org/10.1016/j.ejca.2019.04.001)]
- 4 Silveira CEG, Carcano C, Mauad EC, *et al.* Cell phone usefulness to improve the skin cancer screening: Preliminary results and critical analysis of mobile App development. *Rural and Remote Health*, 2019, 19(1): 4895. [doi: [10.22605/RRH4895](https://doi.org/10.22605/RRH4895)]
- 5 Gilling S, Mortz CG, Vestergaard T. Patient satisfaction and expectations regarding mobile teledermoscopy in general practice for diagnosis of non-melanoma skin cancer and malignant melanoma. *Acta Dermato-venereologica*, 2020, 100(8): adv00117. [doi: [10.2340/00015555-3459](https://doi.org/10.2340/00015555-3459)]
- 6 Rigel DS, Russak J, Friedman R. The evolution of melanoma diagnosis: 25 years beyond the ABCDs. *CA: A Cancer Journal for Clinicians*, 2010, 60(5): 301–316. [doi: [10.3322/caac.20074](https://doi.org/10.3322/caac.20074)]
- 7 Hekler A, Utikal JS, Enk AH, *et al.* Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. *European Journal of Cancer*, 2019, 118: 91–96. [doi: [10.1016/j.ejca.2019.06.012](https://doi.org/10.1016/j.ejca.2019.06.012)]
- 8 Gutman D, Codella NCF, Celebi E, *et al.* Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC). arXiv:1605.01397, 2016.
- 9 Codella NCF, Gutman D, Celebi ME, *et al.* Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). *Proceedings of the 15th IEEE International Symposium on Biomedical Imaging (ISBI 2018)*. Washington: IEEE, 2018. 168–172.
- 10 Della N, Rotemberg V, Tschandl P, *et al.* Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC). arXiv:1902.03368, 2019.
- 11 Rotemberg V, Kurtansky N, Betz-Stablein B, *et al.* A patient-centric dataset of images and metadata for identifying melanomas using clinical context. *Scientific Data*, 2021, 8(1): 34. [doi: [10.1038/s41597-021-00815-z](https://doi.org/10.1038/s41597-021-00815-z)]
- 12 Kurtansky NR, D'Alessandro BM, Gillis MC, *et al.* The SLICE-3D dataset: 400 000 skin lesion image crops extracted from 3D TBP for skin cancer detection. *Scientific Data*, 2024, 11(1): 884. [doi: [10.1038/s41597-024-03743-w](https://doi.org/10.1038/s41597-024-03743-w)]
- 13 Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermoscopic images of common pigmented skin lesions. *Scientific Data*, 2018, 5(1): 180161. [doi: [10.1038/sdata.2018.161](https://doi.org/10.1038/sdata.2018.161)]
- 14 Mendonça T, Ferreira PM, Marques JS, *et al.* PH<sup>2</sup>—A dermoscopic image database for research and benchmarking. *Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Osaka: IEEE, 2013. 5437–5440.
- 15 Bajwa MN, Muta K, Malik MI, *et al.* Computer-aided diagnosis of skin diseases using deep neural networks. *Applied Sciences*, 2020, 10(7): 2488. [doi: [10.3390/app10072488](https://doi.org/10.3390/app10072488)]
- 16 Pacheco AGC, Lima GR, Salomão AS, *et al.* PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. *Data in Brief*, 2020, 32: 106221. [doi: [10.1016/j.dib.2020.106221](https://doi.org/10.1016/j.dib.2020.106221)]
- 17 Kriegsmann K, Lobers F, Zgorzelski C, *et al.* Deep learning for the detection of anatomical tissue structures and neoplasms of the skin on scanned histopathological tissue sections. *Frontiers in Oncology*, 2022, 12: 1022967. [doi: [10.3389/fonc.2022.1022967](https://doi.org/10.3389/fonc.2022.1022967)]
- 18 Kato J, Horimoto K, Sato S, *et al.* Dermoscopy of melanoma and non-melanoma skin cancers. *Frontiers in Medicine*, 2019, 6: 180. [doi: [10.3389/fmed.2019.00180](https://doi.org/10.3389/fmed.2019.00180)]
- 19 LeCun Y, Bottou L, Bengio Y, *et al.* Gradient-based learning applied to document recognition. *Proceedings of*

- the IEEE, 1998, 86(11): 2278–2324. [doi: [10.1109/5.726791](https://doi.org/10.1109/5.726791)]
- 20 Gururaj HL, Manju N, Nagarjun A, *et al.* DeepSkin: A deep learning approach for skin cancer classification. *IEEE Access*, 2023, 11: 50205–50214. [doi: [10.1109/ACCESS.2023.3274848](https://doi.org/10.1109/ACCESS.2023.3274848)]
- 21 Zhang YL, Chen JQ, Wang K, *et al.* ECL: Class-enhancement contrastive learning for long-tailed skin lesion classification. *Proceedings of the 26th International Conference on Medical Image Computing and Computer Assisted Intervention*. Vancouver: Springer, 2023. 244–254.
- 22 Ali MU, Khalid M, Alshambari H, *et al.* Enhancing skin lesion detection: A multistage multiclass convolutional neural network-based framework. *Bioengineering*, 2023, 10(12): 1430. [doi: [10.3390/bioengineering10121430](https://doi.org/10.3390/bioengineering10121430)]
- 23 Cheng HB, Lian J, Jiao WZ. Enhanced MobileNet for skin cancer image classification with fused spatial channel attention mechanism. *Scientific Reports*, 2024, 14(1): 28850. [doi: [10.1038/s41598-024-80087-w](https://doi.org/10.1038/s41598-024-80087-w)]
- 24 Musthafa MM, R MT, V VK, *et al.* Enhanced skin cancer diagnosis using optimized CNN architecture and checkpoints for automated dermatological lesion classification. *BMC Medical Imaging*, 2024, 24: 201. [doi: [10.1186/s12880-024-01356-8](https://doi.org/10.1186/s12880-024-01356-8)]
- 25 Farea E, Saleh RAA, AbuAlkebash H, *et al.* A hybrid deep learning skin cancer prediction framework. *Engineering Science and Technology, an International Journal*, 2024, 57: 101818. [doi: [10.1016/j.jestch.2024.101818](https://doi.org/10.1016/j.jestch.2024.101818)]
- 26 Mehmood A, Gulzar Y, Ilyas QM, *et al.* SBXception: A shallower and broader xception architecture for efficient classification of skin lesions. *Cancers*, 2023, 15(14): 3604. [doi: [10.3390/cancers15143604](https://doi.org/10.3390/cancers15143604)]
- 27 Suleman M, Ullah F, Aldehim G, *et al.* Smart MobiNet: A deep learning approach for accurate skin cancer diagnosis. *Computers, Materials and Continua*, 2023, 77(3): 3533–3549. [doi: [10.32604/cmc.2023.042365](https://doi.org/10.32604/cmc.2023.042365)]
- 28 Shinde RK, Alam MS, Hossain MB, *et al.* Squeeze-mnet: Precise skin cancer detection model for low computing IOT devices using transfer learning. *Cancers*, 2022, 15(1): 12. [doi: [10.3390/cancers15010012](https://doi.org/10.3390/cancers15010012)]
- 29 Medhat S, Abdel-Galil H, Aboutabl AE, *et al.* Iterative magnitude pruning-based light-version of AlexNet for skin cancer classification. *Neural Computing and Applications*, 2024, 36(3): 1413–1428. [doi: [10.1007/s00521-023-09111-w](https://doi.org/10.1007/s00521-023-09111-w)]
- 30 Hossain MM, Hossain MM, Arefin MB, *et al.* Combining state-of-the-art pre-trained deep learning models: A noble approach for skin cancer detection using max voting ensemble. *Diagnostics*, 2023, 14(1): 89. [doi: [10.3390/diagnostics14010089](https://doi.org/10.3390/diagnostics14010089)]
- 31 Thwin SM, Park HS. Skin lesion classification using a deep ensemble model. *Applied Sciences*, 2024, 14(13): 5599. [doi: [10.3390/app14135599](https://doi.org/10.3390/app14135599)]
- 32 Chatterjee S, Gil JM, Byun YC. Early detection of multiclass skin lesions using transfer learning-based IncepX-ensemble model. *IEEE Access*, 2024, 12: 113677–113693. [doi: [10.1109/ACCESS.2024.3432904](https://doi.org/10.1109/ACCESS.2024.3432904)]
- 33 Ichim L, Mitrica RI, Serghei MO, *et al.* Detection of malignant skin lesions based on decision fusion of ensembles of neural networks. *Cancers*, 2023, 15(20): 4946. [doi: [10.3390/cancers15204946](https://doi.org/10.3390/cancers15204946)]
- 34 Ogundokun RO, Li AM, Babatunde RS, *et al.* Enhancing skin cancer detection and classification in dermoscopic images through concatenated MobileNetV2 and xception models. *Bioengineering*, 2023, 10(8): 979. [doi: [10.3390/bioengineering10080979](https://doi.org/10.3390/bioengineering10080979)]
- 35 Dorathi Jayaseeli JD, Briskilal J, Fancy C, *et al.* An intelligent framework for skin cancer detection and classification using fusion of squeeze-excitation-densenet with metaheuristic-driven ensemble deep learning models. *Scientific Reports*, 2025, 15(1): 7425. [doi: [10.1038/s41598-025-92293-1](https://doi.org/10.1038/s41598-025-92293-1)]
- 36 Behara K, Bhero E, Agee JT. Grid-based structural and dimensional skin cancer classification with self-featured optimized explainable deep convolutional neural networks. *International Journal of Molecular Sciences*, 2024, 25(3): 1546. [doi: [10.3390/ijms25031546](https://doi.org/10.3390/ijms25031546)]
- 37 Metta C, Beretta A, Guidotti R, *et al.* Advancing dermatological diagnostics: Interpretable AI for enhanced skin lesion classification. *Diagnostics*, 2024, 14(7): 753. [doi: [10.3390/diagnostics14070753](https://doi.org/10.3390/diagnostics14070753)]
- 38 Abdelhafeez A, Mohamed HK, Maher A, *et al.* A novel approach toward skin cancer classification through fused deep features and neutrosophic environment. *Frontiers in Public Health*, 2023, 11: 1123581. [doi: [10.3389/fpubh.2023.1123581](https://doi.org/10.3389/fpubh.2023.1123581)]
- 39 Olayah F, Senan EM, Ahmed IA, *et al.* AI techniques of dermoscopy image analysis for the early detection of skin lesions based on combined CNN features. *Diagnostics*, 2023, 13(7): 1314. [doi: [10.3390/diagnostics13071314](https://doi.org/10.3390/diagnostics13071314)]
- 40 Ahmed IA, Senan EM, Shatnawi HSA, *et al.* Multi-models of analyzing dermoscopy images for early detection of multi-class skin lesions based on fused features. *Processes*, 2023, 11(3): 910. [doi: [10.3390/pr11030910](https://doi.org/10.3390/pr11030910)]

- 41 Kumar S, Nath VK, Hazarika D. Blend of deep features and binary tree growth algorithm for skin lesion classification. *Symmetry*, 2023, 15(12): 2213. [doi: [10.3390/sym15122213](https://doi.org/10.3390/sym15122213)]
- 42 Zafar M, Amin J, Sharif M, *et al.* DeepLabv3+-based segmentation and best features selection using slime mould algorithm for multi-class skin lesion classification. *Mathematics*, 2023, 11(2): 364. [doi: [10.3390/math11020364](https://doi.org/10.3390/math11020364)]
- 43 Saleh N, Hassan MA, Salaheldin AM. Skin cancer classification based on an optimized convolutional neural network and multicriteria decision-making. *Scientific Reports*, 2024, 14(1): 17323. [doi: [10.1038/s41598-024-67424-9](https://doi.org/10.1038/s41598-024-67424-9)]
- 44 Xin C, Liu ZF, Zhao KY, *et al.* An improved Transformer network for skin cancer classification. *Computers in Biology and Medicine*, 2022, 149: 105939. [doi: [10.1016/j.combiomed.2022.105939](https://doi.org/10.1016/j.combiomed.2022.105939)]
- 45 Dosovitskiy A, Beyer L, Kolesnikov A, *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. *Proceedings of the 9th International Conference on Learning Representations*. Vienna: OpenReview.net, 2021.
- 46 Zhang YJ, Hu YF, Li K, *et al.* Exploring the influence of Transformer-based multimodal modeling on clinicians' diagnosis of skin diseases: A quantitative analysis. *Digital Health*, 2024, 10: 20552076241257087.
- 47 Gallazzi M, Biavaschi S, Bulgheroni A, *et al.* A large dataset to enhance skin cancer classification with Transformer-based deep neural networks. *IEEE Access*, 2024, 12: 109544–109559. [doi: [10.1109/ACCESS.2024.3439365](https://doi.org/10.1109/ACCESS.2024.3439365)]
- 48 Flodorf C, Engelker J, Keller I, *et al.* Skin cancer detection utilizing deep learning: Classification of skin lesion images using a vision Transformer. *arXiv:2407.18554*, 2024.
- 49 Yang G, Luo SH, Greer P. A novel vision Transformer model for skin cancer classification. *Neural Processing Letters*, 2023, 55(7): 9335–9351. [doi: [10.1007/s11063-023-11204-5](https://doi.org/10.1007/s11063-023-11204-5)]
- 50 Aladhah S, Alsanea M, Aloraini M, *et al.* An effective skin cancer classification mechanism via medical vision Transformer. *Sensors*, 2022, 22(11): 4008. [doi: [10.3390/s22114008](https://doi.org/10.3390/s22114008)]
- 51 Himel GMS, Islam MM, Al-Aff KA, *et al.* Skin cancer segmentation and classification using vision Transformer for automatic analysis in dermatoscopy-based noninvasive digital system. *International Journal of Biomedical Imaging*, 2024, 2024(1): 3022192.
- 52 Remya S, Anjali T, Sugumaran V. A novel transfer learning framework for multimodal skin lesion analysis. *IEEE Access*, 2024, 12: 50738–50754. [doi: [10.1109/ACCESS.2024.3385340](https://doi.org/10.1109/ACCESS.2024.3385340)]
- 53 Wang K, Qiu LW, Zhang YL, *et al.* I2M2Net: Inter/intra-modal feature masking self-distillation for incomplete multimodal skin lesion diagnosis. *Proceedings of the 2025 International Conference on Medical Image Computing and Computer Assisted Intervention*. Marrakesh: Springer, 2025. 3–13.
- 54 Gamage L, Isuranga U, Meedeniya D, *et al.* Melanoma skin cancer identification with explainability utilizing mask guided technique. *Electronics*, 2024, 13(4): 680. [doi: [10.3390/electronics13040680](https://doi.org/10.3390/electronics13040680)]
- 55 Pacal I, Alaftekin M, Zengul FD. Enhancing skin cancer diagnosis using swin Transformer with hybrid shifted window-based multi-head self-attention and SwiGLU-based MLP. *Journal of Imaging Informatics in Medicine*, 2024, 37(6): 3174–3192. [doi: [10.1007/s10278-024-01140-8](https://doi.org/10.1007/s10278-024-01140-8)]
- 56 Yadav DP, Sharma B, Chauhan S, *et al.* Dual scale light weight cross attention Transformer for skin lesion classification. *PLoS One*, 2024, 19(12): e0312598. [doi: [10.1371/journal.pone.0312598](https://doi.org/10.1371/journal.pone.0312598)]
- 57 Pacal I, Ozdemir B, Zeynalov J, *et al.* A novel CNN-ViT-based deep learning model for early skin cancer diagnosis. *Biomedical Signal Processing and Control*, 2025, 104: 107627. [doi: [10.1016/j.bspc.2025.107627](https://doi.org/10.1016/j.bspc.2025.107627)]
- 58 Laouarem A, Kara-Mohamed C, Bourennane EB, *et al.* JILDYA-Net: An efficient lightweight multi-class classification architecture for skin lesions. *International Journal of Imaging Systems and Technology*, 2025, 35(3): e70102. [doi: [10.1002/ima.70102](https://doi.org/10.1002/ima.70102)]
- 59 Goodfellow I, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial networks. *Communications of the ACM*, 2020, 63(11): 139–144. [doi: [10.1145/3422622](https://doi.org/10.1145/3422622)]
- 60 Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. *Proceedings of the 4th International Conference on Learning Representations*. San Juan: OpenReview.net, 2016.
- 61 Behara K, Bhero E, Agee JT. Skin lesion synthesis and classification using an improved DCGAN classifier. *Diagnostics*, 2023, 13(16): 2635. [doi: [10.3390/diagnostics13162635](https://doi.org/10.3390/diagnostics13162635)]
- 62 Saeed M, Naseer A, Masood H, *et al.* The power of generative AI to augment for enhanced skin cancer classification: A deep learning approach. *IEEE Access*,

- 2023, 11: 130330–130344. [doi: [10.1109/ACCESS.2023.3332628](https://doi.org/10.1109/ACCESS.2023.3332628)]
- 63 Lee H, Cho S, Song J, *et al.* An enhanced approach using AGS network for skin cancer classification. *Sensors*, 2025, 25(2): 394. [doi: [10.3390/s25020394](https://doi.org/10.3390/s25020394)]
- 64 Pérez E, Ventura S. Progressive growing of generative adversarial networks for improving data augmentation and skin cancer diagnosis. *Artificial Intelligence in Medicine*, 2023, 141: 102556. [doi: [10.1016/j.artmed.2023.102556](https://doi.org/10.1016/j.artmed.2023.102556)]
- 65 Gowthami S, Harikumar R. Residual generative adversarial adaptation network for the classification of melanoma. *International Journal of Computers Communications & Control*, 2023, 18(6). [doi: [10.15837/ijccc.2023.6.5274](https://doi.org/10.15837/ijccc.2023.6.5274)]
- 66 Su QC, Hamed HNA, Isa MA, *et al.* A GAN-based data augmentation method for imbalanced multi-class skin lesion classification. *IEEE Access*, 2024, 12: 16498–16513. [doi: [10.1109/ACCESS.2024.3360215](https://doi.org/10.1109/ACCESS.2024.3360215)]
- 67 Ahmad B, Jun S, Palade V, *et al.* Improving skin cancer classification using heavy-tailed student t-distribution in generative adversarial networks (TED-GAN). *Diagnostics*, 2021, 11(11): 2147. [doi: [10.3390/diagnostics11112147](https://doi.org/10.3390/diagnostics11112147)]
- 68 Nie YL, Sommella P, Carratù M, *et al.* A deep CNN Transformer hybrid model for skin lesion classification of dermoscopic images using focal loss. *Diagnostics*, 2022, 13(1): 72. [doi: [10.3390/diagnostics13010072](https://doi.org/10.3390/diagnostics13010072)]
- 69 Mateen M, Hayat S, Arshad F, *et al.* Hybrid deep learning framework for melanoma diagnosis using dermoscopic medical images. *Diagnostics*, 2024, 14(19): 2242. [doi: [10.3390/diagnostics14192242](https://doi.org/10.3390/diagnostics14192242)]
- 70 Karthik R, Menaka R, Atre S, *et al.* A hybrid deep learning approach for skin cancer classification using swin Transformer and dense group shuffle non-local attention network. *IEEE Access*, 2024, 12: 158040–158051. [doi: [10.1109/ACCESS.2024.3485507](https://doi.org/10.1109/ACCESS.2024.3485507)]
- 71 Ali A, Shahbaz H, Damaševičius R. xCViT: Improved vision Transformer network with fusion of CNN and xception for skin disease recognition with explainable AI. *Computers, Materials and Continua*, 2025, 83(1): 1367–1398. [doi: [10.32604/cmc.2025.059301](https://doi.org/10.32604/cmc.2025.059301)]
- 72 Abbas Q, Daadaa Y, Rashid U, *et al.* Assist-dermo: A lightweight separable vision Transformer model for multiclass skin lesion classification. *Diagnostics*, 2023, 13(15): 2531. [doi: [10.3390/diagnostics13152531](https://doi.org/10.3390/diagnostics13152531)]
- 73 Rezaee K, Zadeh HG. Self-attention Transformer unit-based deep learning framework for skin lesions classification in smart healthcare. *Discover Applied Sciences*, 2024, 6(1): 3. [doi: [10.1007/s42452-024-05655-1](https://doi.org/10.1007/s42452-024-05655-1)]
- 74 Islam N, Hasib KM, Joti FA, *et al.* Leveraging knowledge distillation for lightweight skin cancer classification: Balancing accuracy and computational efficiency. arXiv:2406.17051, 2024.
- 75 Foltz EA, Ludzik J, Leachman S, *et al.* Revolutionizing skin cancer triage: The role of patient-initiated teledermoscopy in remote diagnosis. *Cancers*, 2024, 16(14): 2565. [doi: [10.3390/cancers16142565](https://doi.org/10.3390/cancers16142565)]
- 76 Lee C, Witkowski A, Żychowska M, *et al.* The role of mobile teledermoscopy in skin cancer triage and management during the COVID-19 pandemic. *Indian Journal of Dermatology, Venereology and Leprology*, 2023, 89(3): 347–352.
- 77 Gregoor AMS, Sangers TE, Eekhof JAH, *et al.* Artificial intelligence in mobile health for skin cancer diagnostics at home (AIM HIGH): A pilot feasibility study. *eClinicalMedicine*, 2023, 60: 102019. [doi: [10.1016/j.eclinm.2023.102019](https://doi.org/10.1016/j.eclinm.2023.102019)]
- 78 Azeem M, Kiani K, Mansouri T, *et al.* SkinLesNet: Classification of skin lesions and detection of melanoma cancer using a novel multi-layer deep convolutional neural network. *Cancers*, 2024, 16(1): 108.
- 79 Wang H, Qi QQ, Sun WJ, *et al.* Classification of clinical skin lesions with double-branch networks. *Frontiers in Medicine*, 2023, 10: 1114362. [doi: [10.3389/fmed.2023.1114362](https://doi.org/10.3389/fmed.2023.1114362)]
- 80 Chen TL, Liu Q, Yang J. Few-shot classification with multiscale feature fusion for clinical skin disease diagnosis. *Clinical, Cosmetic and Investigational Dermatology*, 2024, 17: 1007–1026. [doi: [10.2147/CCID.S458255](https://doi.org/10.2147/CCID.S458255)]
- 81 Teymouri S, Larki AA, Mohammadi M, *et al.* Hybrid vision Transformer-boosting model for skin lesion classification on SLICE-3D. *Proceedings of the 12th RSI International Conference on Robotics and Mechatronics (ICRoM)*. Tehran: IEEE, 2024. 586–592.
- 82 Wang K, Shan NY, Gouk H, *et al.* Skin malignancy classification using patients' skin images and meta-data: Multimodal fusion for improving fairness. *Proceedings of the 7th International Conference on Medical Imaging with Deep Learning*. Paris: PMLR, 2024. 1670–1686.
- 83 Yan SY, Yu Z, Primiero C, *et al.* A multimodal vision foundation model for clinical dermatology. *Nature Medicine*, 2025, 31(8): 2691–2702. [doi: [10.1038/s41591-025-03747-y](https://doi.org/10.1038/s41591-025-03747-y)]
- 84 Zhou YJ, Liu W, Gao Y, *et al.* A novel multi-task model imitating dermatologists for accurate differential diagnosis of skin diseases in clinical images. *Proceedings of the 26th*

- International Conference on Medical Image Computing and Computer Assisted Intervention. Vancouver: Springer, 2023. 202–212.
- 85 Wang J, Chung Y, Ding Z, *et al.* From majority to minority: A diffusion-based augmentation for underrepresented groups in skin lesion analysis. Proceedings of the 2025 International Conference on Medical Image Computing and Computer Assisted Intervention. Marrakesh: Springer, 2025. 14–23.
- 86 Gottfrois P, Gröger F, Andriamboloniaina FH, *et al.* PASSION for dermatology: Bridging the diversity gap with pigmented skin images from sub-Saharan Africa. Proceedings of the 27th International Conference on Medical Image Computing and Computer Assisted Intervention. Marrakesh: Springer, 2024. 703–712.
- 87 Kim AS, Saharkhiz N, Sizikova E, *et al.* S-SYNTH: Knowledge-based, synthetic generation of skin images. Proceedings of the 27th International Conference on Medical Image Computing and Computer Assisted Intervention. Marrakesh: Springer, 2024. 734–744.
- 88 Kabir MR, Borshon RH, Wasi MK, *et al.* Skin cancer detection using lightweight model souping and ensembling knowledge distillation for memory-constrained devices. Intelligence-based Medicine, 2024, 10: 100176. [doi: [10.1016/j.ibmed.2024.100176](https://doi.org/10.1016/j.ibmed.2024.100176)]
- 89 Watt T, Chrysoulas C, Barclay PJ. Moving healthcare AI support systems for visually detectable diseases to constrained devices. Applied Sciences, 2024, 14(24): 11474. [doi: [10.3390/app142411474](https://doi.org/10.3390/app142411474)]
- 90 Carse J, Süveges T, Chin G, *et al.* Classifying real-world macroscopic images in the primary-secondary care interface using transfer learning: Implications for development of artificial intelligence solutions using nondermoscopic images. Clinical and Experimental Dermatology, 2024, 49(7): 699–706. [doi: [10.1093/ced/llad400](https://doi.org/10.1093/ced/llad400)]
- 91 Somfai E, Baffy B, Fenech K, *et al.* Handling dataset dependence with model ensembles for skin lesion classification from dermoscopic and clinical images. International Journal of Imaging Systems and Technology, 2023, 33(2): 556–571. [doi: [10.1002/ima.22827](https://doi.org/10.1002/ima.22827)]
- 92 Santos ARF, Aires KRT, Veras RMS. Aspects of lighting and color in classifying malignant skin cancer with deep learning. Applied sciences, 2024, 14(8): 3297. [doi: [10.3390/app14083297](https://doi.org/10.3390/app14083297)]
- 93 Al Mahmud A, Azam S, Khan IU, *et al.* SkinNet-14: A deep learning framework for accurate skin cancer classification using low-resolution dermoscopy images with optimized training time. Neural Computing and Applications, 2024, 36(30): 18935–18959. [doi: [10.1007/s00521-024-10225-y](https://doi.org/10.1007/s00521-024-10225-y)]
- 94 Li M, Abe M, Nakano S, *et al.* Deep learning approach to classify cutaneous melanoma in a whole slide image. Cancers, 2023, 15(6): 1907. [doi: [10.3390/cancers15061907](https://doi.org/10.3390/cancers15061907)]
- 95 Frago-Garcia M, Wilm F, Bertram CA, *et al.* Automated diagnosis of 7 canine skin tumors using machine learning on H&E-stained whole slide images. Veterinary Pathology, 2023, 60(6): 865–875. [doi: [10.1177/03009858231189205](https://doi.org/10.1177/03009858231189205)]
- 96 Chaurasia AK, Toohey PW, Harris HC, *et al.* Multi-resolution vision Transformer model for skin cancer subtype classification using histopathology slides. <https://www.medrxiv.org/content/10.1101/2025.01.28.25321234v1.full.pdf>. [2025-01-28]. [doi: [10.1101/2025.01.28.25321234](https://doi.org/10.1101/2025.01.28.25321234)]
- 97 Feng M, Xu KL, Wu NH, *et al.* Trusted multi-scale classification framework for whole slide image. Biomedical Signal Processing and Control, 2024, 89: 105790. [doi: [10.1016/j.bspc.2023.105790](https://doi.org/10.1016/j.bspc.2023.105790)]
- 98 Shi ZJ, Zhu JY, Yu LH, *et al.* A two-stage end-to-end deep learning framework for pathologic examination in skin tumor diagnosis. The American Journal of Pathology, 2023, 193(6): 769–777. [doi: [10.1016/j.ajpath.2023.02.008](https://doi.org/10.1016/j.ajpath.2023.02.008)]
- 99 Nejat P, Alsaafin A, Alabtah G, *et al.* Creating an atlas of normal tissue for pruning WSI patching through anomaly detection. Scientific Reports, 2024, 14(1): 3932. [doi: [10.1038/s41598-024-54489-9](https://doi.org/10.1038/s41598-024-54489-9)]
- 100 Geijs DJ, Dooper S, Aswolinskiy W, *et al.* Detection and subtyping of basal cell carcinoma in whole-slide histopathology using weakly-supervised learning. Medical Image Analysis, 2024, 93: 103063. [doi: [10.1016/j.media.2023.103063](https://doi.org/10.1016/j.media.2023.103063)]
- 101 Björkman J, Lagerroth S, Siarov J, *et al.* Enhancing basal cell carcinoma classification in preoperative biopsies via transfer learning with weakly supervised graph Transformers. BMC Medical Imaging, 2025, 25(1): 166. [doi: [10.1186/s12880-025-01710-4](https://doi.org/10.1186/s12880-025-01710-4)]
- 102 Tran M, Schmidle P, Guo RR, *et al.* Generating dermatopathology reports from gigapixel whole slide images with HistoGPT. Nature Communications, 2025, 16(1): 4886. [doi: [10.1038/s41467-025-60014-x](https://doi.org/10.1038/s41467-025-60014-x)]
- 103 Meseguer P, Del Amor R, Naranjo V. MICIL: Multiple-instance class-incremental learning for skin cancer whole slide images. Artificial Intelligence in Medicine, 2024, 152: 102870. [doi: [10.1016/j.artmed.2024.102870](https://doi.org/10.1016/j.artmed.2024.102870)]

(校对责编: 李慧鑫)