

深度学习与多模态大语言模型 在笑线分类任务中的应用初探

孟笑菲¹ 龚卓弘² 万耀文³ 李沛达³ 胡琪琪¹ 邱龙诗语⁴ 刘恒毅⁴ 谢伟丽¹

¹哈尔滨医科大学附属第一医院口腔修复科, 哈尔滨 150001; ²香港大学牙医学院修复齿科, 香港 999077; ³中山大学计算机学院, 广州 510006; ⁴中山大学附属口腔医院, 光华口腔医学院, 广东省口腔医学重点实验室, 广东省口腔疾病临床医学研究中心, 广州 510055

通信作者: 谢伟丽, Email: xiewl@hrbmu.edu.cn

【摘要】目的 口腔微笑美学评估是临床治疗规划的关键环节, 笑线作为关键美学指标, 其精准分类对优化美学修复重建治疗方案具有重要意义。然而, 笑线分类需要对唇、龈和齿之间的复杂关系准确判断, 医师分析的主观性可能造成误诊。本研究旨在以笑线分类任务为口腔微笑美学智能化初探, 对比卷积神经网络(CNN)、大语言模型(LLM)及不同层次医师在口腔微笑线分类任务中的表现。**方法** 以公开高质量人脸数据集FFHQ为基础, 经图像预处理与标准化标注后构建含1000张样本的微笑图像标注数据集, 标注内容为高、中、低笑线3种类型。采用7种经典CNN模型(VGG16、ResNet34等)与5种代表性多模态LLM(Qwen系列、LLaVA-1.5-7B)进行训练、验证和测试, 通过准确率、精确率、召回率及 F_1 分数比较模型性能, 并与不同层级临床医师的评估结果进行对比。**结果** 在7种常用CNN模型中, ResNet152模型总体表现最优, 分类准确率达83.30%, 显著优于其他CNN模型及多模态LLM; 高级口腔医师分类准确率为83.00%, 与ResNet152性能接近。注意力热力图显示ResNet152模型关注区域与医师相似。**结论** CNN模型在笑线分类任务中具备更高的临床应用潜力, 可达到专家水平; LLM的医学图像精细分类能力仍需优化。本研究为口腔美学智能评估系统的开发提供了实验依据与技术参考。

【关键词】 微笑美学; 笑线分类; 深度学习; 大语言模型

引用著录格式: 孟笑菲, 龚卓弘, 万耀文, 等. 深度学习与多模态大语言模型在笑线分类任务中的应用初探[JOL]. 中华口腔医学研究杂志(电子版), 2026, 20(1): 17-24.

DOI: 10.3877/cma.j.issn.1674-1366.2026.01.003

Preliminary exploration on the application of the classification of dental smile lines via deep learning and multimodal large language model

Meng Xiaofei¹, Gong Zhuohong², Wan Yaowen³, Li Peida³, Hu Qiqi¹, Qiu Longshiyu⁴, Liu Hengyi⁴, Xie Weili¹

¹Department of Prosthodontics, The First Affiliated Hospital of Harbin Medical University, Harbin 150001, China; ²Restorative Dental Sciences, Faculty of Dentistry, The University of Hong Kong, Hong Kong 999077, China; ³School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006, China; ⁴Hospital of Stomatology, Guanghua School of Stomatology, Sun Yat-sen University, Guangdong Provincial Key Laboratory of Stomatology, Guangdong Provincial Clinical Research Center of Oral Diseases, Guangzhou 510055, China

Corresponding author: Xie Weili, Email: xiewl@hrbmu.edu.cn

【Abstract】Objective Smile aesthetic evaluation is a critical component of clinical treatment planning. As a key aesthetic indicator, the precise classification of the smile line is crucial for optimizing restorative and reconstructive treatment plans. However, classifying smile line requires accurate assessment of complex relationships among the lips, gingiva and teeth, and analysis by dentists involves a degree of subjectivity and chances for misdiagnosis. This study aimed to investigate smile line classification

by comparing the performance of convolutional neural networks (CNNs) and large language models (LLMs), as well as clinicians of varying expertise levels, in this task. **Methods** Based on the publicly available high-quality FFHQ facial dataset, a smile image annotation dataset comprising 1 000 samples was constructed following image preprocessing and standardized annotations of three types: high, medium and low smile line. Seven classic CNN models (VGG16, ResNet34, etc.) and five representative multimodal LLMs (Qwen series, LLaVA 1.5-7B) were employed for training, validation, and testing. Model performance was evaluated using accuracy, precision, recall, and F_1 scores, and compared against assessments made by clinicians of different seniority levels. **Results** Among the seven commonly used CNN models, the ResNet152 model demonstrated optimal overall performance, achieving a mean classification accuracy of 83.30%, which significantly outperformed other CNN models and multimodal LLMs. Senior dentists achieved a classification accuracy of 83.00%, comparable to the performance of ResNet152. Heatmaps demonstrate similar attention regions between ResNet152 and dental practitioners. **Conclusions** CNN models demonstrated substantial clinical potential in smile line classification tasks, attaining expert-level performance. In contrast, large language models required further optimization for medical image fine-grained classification. This study provided experimental evidence and technical insights for developing intelligent aesthetic assessment systems in dentistry.

【Key words】 Smile esthetics; Classification of smile lines; Deep learning; Large language models

DOI:10.3877/cma.j.issn.1674-1366.2026.01.003

美学评估在口腔临床诊断和治疗中起到至关重要的作用,其中获得美观和谐的微笑是口腔美学评估及美学治疗的终极目标^[1]。微笑美学分析与修复、牙周、种植和正畸等多个口腔医学专科有密切关系,已经成为专业人员客观评价治疗效果的重要指标^[2]。全面的微笑美学分析涉及对面部、嘴唇、牙龈及牙齿等诸多解剖与形态学参数的动静态协同评估,其中,笑线作为微笑美学的关键指标,对美学效果评估具有重要的临床意义^[3]。笑线是指人在自然、放松地微笑时,上唇下缘所处的位置,该分型与口腔美学修复重建诊疗及美学评估有重大影响:对于低笑线类型的人群,牙龈不协调通常不影响微笑美观;而对于中、高笑线类型人群,牙齿及牙龈暴露得越多,美学修复重建的风险越大^[4]。因此,准确的 smile 分类有助于口腔医师制定更合适的美学修复重建治疗方案,带来可预期的治疗效果。

视诊作为识别口腔笑线分类的重要筛查方法,可迅速提供初步分析结果,有助于医师作出进一步的诊疗决策。但是,由于临床医师的诊断标准和经验水平不一,且笑线分类还涉及对上唇、牙龈与牙齿空间位置关系的主观判断,容易出现笑线分类结果不一致。同时,由于缺乏客观数据支持及可视化沟通工具,医师在向患者解释病情和治疗方案时,往往难以达成有效共识,易引发沟通障碍及期望值偏差。这些问题不仅可能造成漏诊、误诊,影响治

疗效果,还会增加医患沟通成本,导致诊疗过程耗时耗力,制约了美学修复整体质量的提升。综上所述,精准且客观的笑线分型、分类对口腔美学评估十分重要。

围绕笑线分类的自动化,已有学者探索应用计算机视觉方法实现自动笑线分型。Lee等^[5]应用OpenCV计算机视觉方法分别分割微笑图像中的上颌牙龈及牙齿区域,并通过牙龈与牙齿面积比值进行笑线分类的自动半定量分析。尽管该研究取得较高的分类准确性,然而其所采用的方法仍为传统的特征提取其表征,尚未采用基于深度学习的方法实现全自动智能化分析。随着人工智能(artificial intelligence, AI)技术的发展,深度学习作为其中一项重要的技术,已被用于医学图像分类^[6]、检测^[7]和分割^[8],并可达到与专家相匹配的准确性^[9]。卷积神经网络(convolutional neural network, CNN)作为广泛应用的深度学习模型,通过卷积层提取空间特征,可以高效识别与分析二维图像,在医学图像分类任务中表现出较高的准确性^[10]。近年来,大语言模型(large language model, LLM)的更新迭代使其具有强大的通用与多模态分析及泛化能力,部分多模态LLM(如Qwen系列、LLaVA-1.5-7B)已具备直接解析图像并结合文本进行推理的能力,为口腔美学智能分析带来了新的可能^[11]。然而,笑线分类涉及对图像中的唇、龈和齿之间的复杂空间关系,对模

型的特征提取及分析能力带来较大的挑战。

为此,本研究旨在构建基于CNN与多模态LLM技术基础的AI模型,探索不同AI模型在摄影图像中进行笑线智能分类任务中的性能,并与各层次临床医师进行比较,探索AI在提供辅助美学评估和治疗规划的临床有效性。

资料与方法

一、数据集

本研究使用公开的高质量人脸数据集FFHQ(Flickr-Faces-High Quality),该数据集包含70 000余张 $1\ 024 \times 1\ 024$ 分辨率的高质量PNG彩色图像,涵盖了不同年龄、种族和图像背景的丰富变化的微笑表情图像,确保数据的多样性和代表性。为保证数据集的质量和一致性,从FFHQ数据集中随机挑选适合进行微笑分析的图像。

1.入选标准:(1)人物的正面微笑;(2)可以清晰分辨嘴唇、牙龈及牙齿。

2.排除标准:(1)图像整体或口腔区域模糊;(2)被食物、手、话筒或其他物品遮挡口腔区域。最终共随机纳入1 000张符合要求的微笑图片,以供后续模型进行训练、验证及测试。从中随机挑选100张作为测试集,用于对比不同模型及各级临床医师的微笑分析性能。

二、图像预处理和标注

由于原始图片范围较大且背景复杂,需要首先对图像进行标准化预处理,减少由于图像背景噪声带来的干扰。使用LabelMe^[12]标注软件(v5.5.0)手动标注微笑区域,以边界框形式标记上下唇之间和两侧口角之间的微笑范围,并应用软件裁剪出微笑区域,完成图像预处理。

由具有临床经验的口腔科医师对预处理后图像进行笑线分类标注,并由1位高级医师进行复核以确保标注一致性。本研究采用Tjan等^[13]提出的标准进行笑线分类,根据微笑时牙齿与牙龈组织的显露量,将笑线划分为高位、中位和低位3种。

1.高位笑线:上唇下缘位于龈缘上方,暴露出全部牙冠及部分牙龈。

2.中位笑线:暴露出全部牙冠但却不暴露牙龈,或上唇少量覆盖上前牙牙冠,露出牙冠面积为75%~100%。

3.低位笑线:上唇下缘覆盖部分上前牙牙冠,露出牙冠面积少于75%。

三、模型的训练和测试

本研究首先使用VGG16、ResNet34、ResNet50、ResNet101、ResNet152、Inception_V3和ResNext101等7种常用CNN模型进行笑线分类任务,对比模型性能并挑选最优CNN模型(图1)。其中,ResNet系列模型为深度残差网络,引入残差模块和瓶颈结构,使得模型可以在增加模型深度的同时有效地学习图像特征。采用7:2:1的比例将标注数据集随机划分为训练集、验证集和测试集。实验在Python3.13平台和PyTorch框架下实施,训练过程中使用了AdamW优化器,训练轮数为30,学习率设定为0.000 1,批量大小为64,权重衰减为0.05。采用迁移学习方法,使用ImageNet上预训练的权重提高模型的特征表示能力。为增强模型鲁棒性并缓解过拟合,训练过程中应用了数据增强策略,包括小幅度的随机缩放、旋转、裁剪及随机的颜色调整与水平翻转,以确保图像中与笑线分类相关的关键区域(如唇、龈、齿)得以保留。

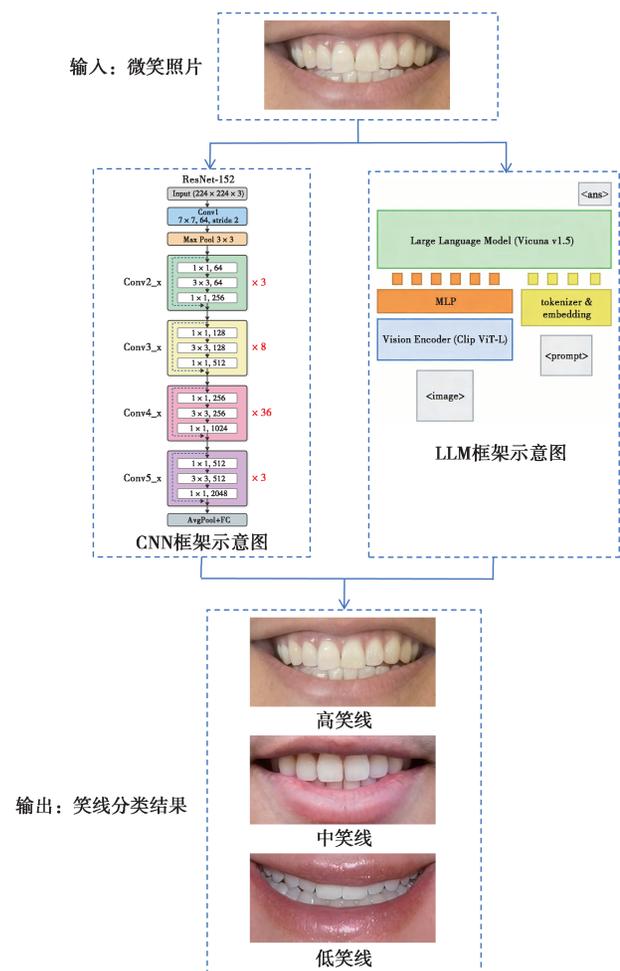


图1 基于卷积神经网络(CNN)和大语言模型(LLM)进行笑线自动分类的模式图

接着,本研究对多款主流预训练多模态大模型开展了基于LoRA(Low-Rank Adaptation)的监督微调实验,对比不同多模态LLM的笑线分析性能,包括Qwen2.5-vl-7B、Qwen2.5-vl-3B、Qwen2-vl-2B、Qwen2-vl-7B和LLaVA-1.5-7B(图1)。为图像设计统一的引导文本构成图文对作为输入,同时通过构造分类导向的监督标签信号引导模型输出类别决策结果。训练阶段固定模型主干网络参数,仅优化低秩适配矩阵,具体将LoRA模块挂载于LLM的注意力层,设置秩为8以平衡训练效率与适配能力;训练过程采用分层学习率调度与交叉熵损失函数等保障收敛。完成训练后,将测试集的图像和引导文本输入到模型中进行测试。

四、模型性能评估和比较

使用准确率、精确率、召回率和 F_1 分数来评估模型的笑线分类性能。

1. 准确率:分类正确的图像数量占总图像数量的比例,反映模型和医师分类的整体准确性。

2. 精确率:预测为某一类笑线类型的图像中,实际属于该类型的比例,衡量模型和医师在每个类别上的分类精度。

3. 召回率:实际属于某一类笑线类型的图像中,被正确预测为该类型的比例,体现模型和医师对每个类别的识别能力。

4. F_1 分数:精确率和召回率的调和平均数,综合评估模型和医师的分类性能,计算公式如下

$$F_1 = 2 \times \frac{\text{精确率} \times \text{召回率}}{\text{精确率} + \text{召回率}}$$

五、人机比赛

邀请不同级别的口腔医师对测试集上的图片进行笑线分类评估,包括2位工作经验超过5年的高级医师、2位工作经验小于5年的初级医师和2位无临床经验的医学生。所有评估者均在评估前接受了统一的标准流程培训,以确保对笑线分类标准理解的一致性。培训内容包括:首先,由同一资深口腔医师讲解Tjan等^[13]提出的笑线分类标准;随后,提供20张未包含于后续数据集的样本图片给不同级别口腔医师进行练习与讨论;最后,评估者独立完成练习后,由资深口腔医师复核并反馈,直至所有评估者对分类标准达成共识。评估过程中,每位医师在统一环境与屏幕设置下,独立进行对测试集中的图像进行笑线分类,并记录标注每张图像所耗费的时间。流程旨在最大限度减少因经验差异

导致的主观偏差,提升医师评估结果之间的可比性及与人机比较的可信度。

六、可视化分析

采用梯度类激活映射(Grad-CAM)定位模型在进行图像分析过程中,进行笑线识别与分类的关键区域,该方法通过模型最后一个卷积层的梯度信息生成热力图,热力图的红色区域代表模型关注的程度^[14]。通过这种方式,可以直观地观察模型关注的图像区域,分析其决策依据,进一步理解模型在图像分类任务中的表现。

七、统计学处理方法

选取7种常用CNN模型和5种常用多模态大模型进行对比实验。CNN模型包括:VGG16、ResNet34、ResNet50、ResNet101、ResNet152、Inception_V3和ResNeXt101;多模态大模型包括:Qwen系列、LLaVA-1.5-7B。所有模型均在相同的训练集和测试集上进行评估,以性能最优的CNN和性能最优的LLM作为目标基准模型,旨在检验其性能是否显著优于其他对比模型,本研究采用配对样本 t 检验进行统计推断,检验水准 $\alpha=0.05$ 。针对数据加载器(dataloader)与数据增强器(transform)引入随机种子控制,每次实验均记录模型在测试集上的准确率、精确率、召回率与 F_1 分数指标,最终通过对每个模型的重复实验的结果取算术平均值,得到各模型在对应指标上的稳定性性能表现。这种重复实验并取均值的方式,能够有效降低随机因素对结果的干扰,更客观地反映模型的实际泛化能力;同时,通过配对 t 检验进一步验证了不同模型间性能差异的统计显著性。

结 果

从公开的高质量人脸数据集FFHQ获取1000张微笑图像数据集,由具有丰富临床经验的口腔科医师对预处理后图像进行笑线分类标注,标注内容为高、中、低笑线3种类型,最终得到高笑线288张、中笑线455张和低笑线257张。其中,测试集中高笑线34张、中笑线51张和低笑线15张。

对7种常见的CNN模型的性能进行了评估,其在测试集上的性能如图2所示,结果表明,ResNet152模型的分类准确率最高,达到83.30%。此外,该模型在精确率、召回率和 F_1 分数等指标上也优于其他模型,表现出最佳的特征提取能力和分类性能。ResNet34、ResNet50、VGG16和ResNet101也表现出较强的分类能力和稳定性,整体性能较

好。相对而言, Inception_V3的性能较弱, 分类准确率较低。以 ResNet152 为基准模型, 采用单尾配对 t 检验比较与其他模型的性能差异, 结果显示 ResNet152 的分类准确率显著优于其他模型, 除了与 VGG16 及 ResNext101 之间的性能差异未达到统计学上的显著性水平。这表明 ResNet152 在本研究的数据集和实验设置下展现出了相对于多数对比模型的稳定优势。

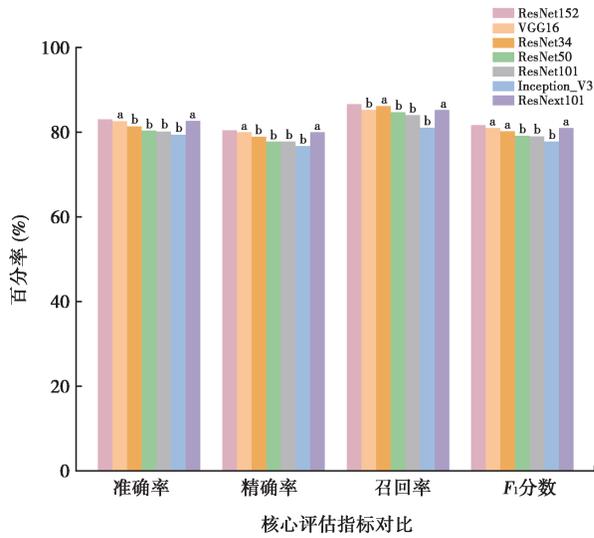


图2 7种卷积神经网络(CNN)模型统计学分析结果 ^a模型性能与 ResNet152 之间的差异无统计学意义 ($P>0.05$); ^b模型性能与 ResNet152 相比差异具有统计学意义 ($P<0.05$)

选择5种具有代表性的多模态 LLM 进行测试, 其在测试集上的性能如图3所示, 结果表明 LLaVA-1.5-7B 模型的分类准确率最高, 达到 62.60%。此外, 该模型在精确率、召回率和 F_1 分数等指标上也优于其他多模态 LLM, 然而均显著低于传统 CNN 模型。以 LLaVA-1.5-7B 为基准模型, 采用单尾配对 t 检验比较与其他模型的性能差异, 结果显示 LLaVA-1.5-7B 的分类准确率显著优于其他模型, 除了与 Qwen2-VL-7B 之间性能差异无统计学意义。这表明在本研究的数据集和实验设置下, LLaVA-1.5-7B 展现出了相对于多数对比模型的稳定优势。

邀请各级口腔医师与最优 CNN 模型(ResNet152)

及最优 LLM(LLaVA-1.5-7B)进行人机比赛, 其在测试集上的性能差异如表1所示。结果表明, ResNet152 模型表现优于其他医师及 LLM。初级医师表现优于高级医师和医学生, 其平均准确率达 85.50%, F_1 分数达到 83.42%; 最优 CNN 模型效率为最优 LLM 的 50 余倍, 是各级口腔医师的 400 倍以上, 总体分类性能与最优 CNN 模型 ResNet152 接近。相比之下, 最优 LLM 在分类性能及耗费时间上没有优势。

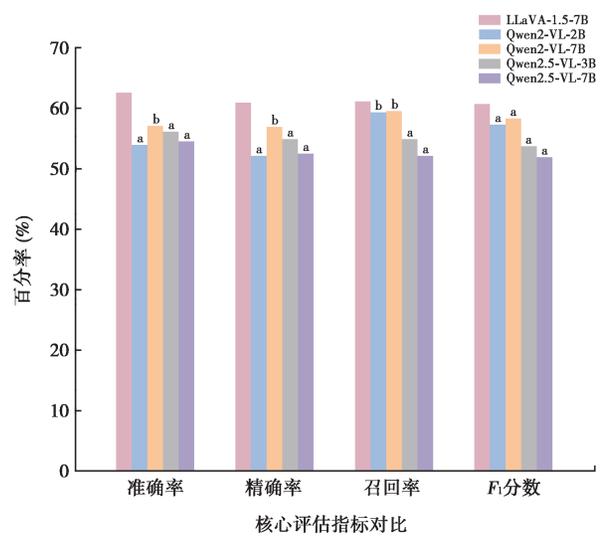


图3 5种大语言模型(LLM)统计学分析结果 ^a模型性能与 LLaVA-1.5-7B 之间的差异无统计学意义 ($P>0.05$); ^b模型性能与 LLaVA-1.5-7B 相比差异具有统计学意义 ($P<0.05$)

各层次最佳口腔医师与最优 CNN 模型 (ResNet152) 及最优 LLM(LLaVA-1.5-7B) 在测试集上的混淆矩阵结果如图4所示。结果显示, ResNet152 模型在识别高笑线时的召回率为 91.18%, 识别中笑线时的召回率为 80.39%, 识别低笑线召回率为 93.33%。LLaVA-1.5-7B LLM 在识别高笑线时召回率为 52.94%, 识别中笑线召回率为 68.63%, 识别低笑线召回率为 53.33%。医学生在中笑线上召回率最优达到 86.27%, 高级医师则在低笑线识别召回率最高达到 86.67%, 而总体性能最佳的初级医师识别高笑线、中笑线及低笑线的召回率分别为 97.06%、78.43% 及 80.00%。

表1 各级医师与最佳卷积神经网络(CNN)和大语言模型(LLM)在测试集上的性能比较

分析者	平均准确率 (%)	平均精确率 (%)	平均召回率 (%)	平均 F_1 分数 (%)	平均时间 (s/张, $\bar{x} \pm s$)	
医师	医学生	83.50	82.55	79.31	78.31	5.060 0 \pm 2.145 2
	初级医师	85.50	83.00	85.13	83.42	4.993 5 \pm 1.987 2
	高级医师	83.00	84.50	80.55	80.35	4.132 8 \pm 1.153 3
最优 CNN	ResNet152	83.30	80.67	86.81	81.84	0.008 9 \pm 0.002 0
最优 LLM	LLaVA-1.5-7B	62.60	61.00	61.20	60.80	0.464 3 \pm 0.017 0

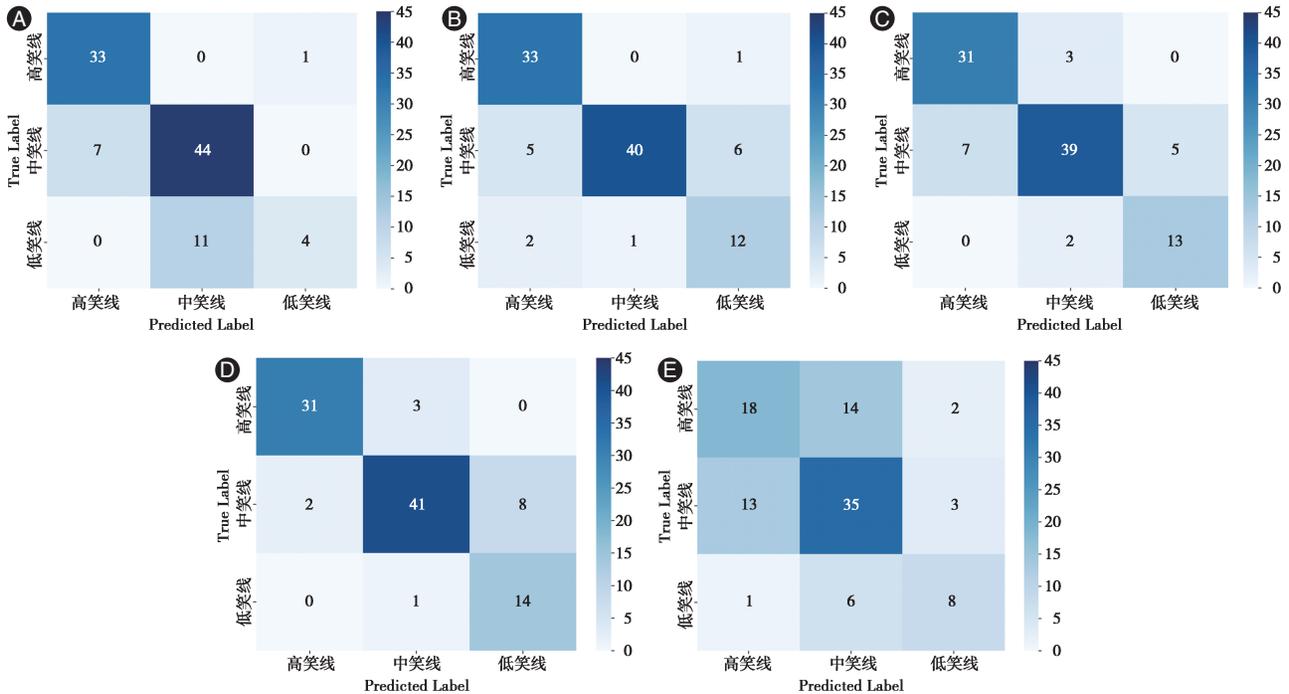


图4 不同层次口腔医师/医学生与人工智能模型进行笑线分类任务的混淆矩阵图 A~C:医学生、初级医师、高级医师混淆矩阵图;D:ResNet152混淆矩阵图;E:LLaVA-1.5-7B混淆矩阵图。

选取分类准确率最高CNN模型 ResNet152,图5展示了基于 ResNet152 模型的注意力热力图,其中红色和黄色代表模型对该区域关注度高,对分类决策的贡献较大;蓝色和绿色表示模型对该区域的关注度低,对分类决策的贡献较小。结果显示, ResNet152 模型在分类微笑图像时,其注意力明显聚焦于上唇及上颌牙龈和牙齿区域,有效忽略了无关背景的干扰。这表明模型能够精准定位并提取与笑线分类任务高度相关的视觉特征,其注意力分布模式与口腔医师在分析临床图像时的视觉关注习惯趋于一致。

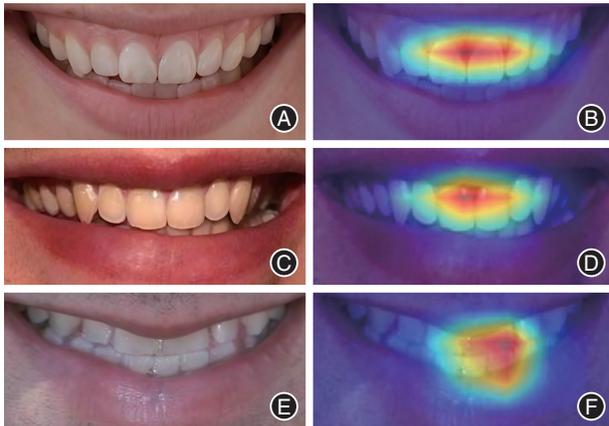


图5 最佳卷积神经网络(CNN)模型(ResNet152网络)进行笑线分类任务的热力图 A~B:高笑线示例图片及热力图;C~D:中笑线示例图片及热力图;E~F:低笑线示例图片及热力图。

讨 论

本研究应用基于CNN与LLM技术原理的AI技术,创新地将其应用于口腔笑线分类任务中,结果显示AI技术可以实现较准确的笑线分类,最优CNN模型的性能优于高级医师,且可以准确捕捉与分类相关的重要图像特征。此外,对口腔笑线智能化分析的初探显示传统CNN模型性能优于LLM,证实传统方法在较为常规的任务中仍具有强鲁棒性。本研究为口腔美学智能评估系统的开发提供了实验依据与技术参考。

一、卷积神经网络与多模态大语言模型在笑线分类任务中的性能比较及原因分析

实验结果表明,在口腔微笑图像分类任务中, ResNet152 模型展现出最佳的分类性能,其准确率达 83.30%。此外,使用 LLM 对同样的微笑图片进行测试, LLaVA-1.5-7B LLM 展现出最佳的分类性能,其准确率为 62.60%,与 CNN 网络相比欠佳。CNN 网络专门设计用于处理具有网格结构的数据,它的卷积层能够自动提取图像的局部特征,如边缘、纹理等,通过池化层进行特征降维,在图像分类^[15]、目标检测^[16]和语义分割^[17]等任务上表现出色。在笑线分类任务中,基于 CNN 的模型能够达到较高的准确率。然而,LoRA 监督微调模型的泛化能力显著

受限,其核心评价指标(准确率、 F_1 分数等)低于传统CNN方法,且与训练集上的表现形成明显落差,呈现出典型的过拟合特征。结合医学分类任务的领域特性,深入分析其过拟合成因主要是:所采用的预训练大模型缺乏医学领域专属预训练过程,其原生知识体系中缺乏目标任务所需的医学专业先验,同时该医学分类任务依赖细粒度的医学特征进行分类判定,再加上数据量不足,这使得模型本身难以完成医学场景下的精准特征提取,难以有效聚焦任务关键的细粒度医学特征,进而过度拟合训练数据中的局部噪声信息。LLM主要是基于文本数据进行训练的,虽然一些多模态LLM也具备一定的图像理解能力,但在图像识别和处理的准确性上通常难以与专业的CNN网络相媲美。因为,其架构和训练方式并非专门针对图像特征的精细提取和分析,与CNN网络相比准确率较低。CNN模型在笑线分类任务中具备更高的临床应用潜力,多模态LLM的图像精细分类能力仍需优化。

二、人机比赛结果分析与临床有效性验证

人机比赛在同一实验设置下对比AI模型与各层次医师之间在同一任务的性能,进而判断AI模型在完成实际任务上是否具有临床有效性。本研究结果显示,ResNet152模型在人机比赛中优于高级医师,并且耗时显著更低,说明AI模型在笑线分类这一任务上已经达到专家性能。AI模型通过卷积等方法,对图像中所有精细的特征进行筛选及特征提取,可以更高效地判断上唇与牙龈、牙齿之间的位置关系;同时,基于公开数据集的大批量标注数据,模型有效学习到与高、中、低笑线分类最相关的特征,注意力热力图也显示模型有效关注与分类最相关的区域。而各层次医师之间也存在较为明显的分类性能差异,特别是医学生对低笑线的分类召回率很低,说明人类医师判断的主观性较大。因此,本研究通过对笑线分类任务的智能化初探,揭示AI模型在口腔美学分析中的客观性、准确性和对微小特征的精确分析潜力,为未来口腔美学智能化一体化评估系统提供证据支撑。

三、人工智能模型在笑线分类任务中的挑战与未来方向

本研究选用公共数据集图像作为分类基础,提升了数据获取的便捷性,为大规模临床应用探索提供了可行路径。本研究也存在一定局限性。首先,FFHQ作为高质量人脸数据集,其图像多为在受控

环境下拍摄的正面肖像,具有较好的光照条件与拍摄角度;然而,真实口腔临床拍摄参数偏差大,饱和度、亮度、清晰度、放大率存在差异且受诊室光照不均、患者头部姿态自然变化的影响,这种采集场景的差异意味着本研究训练的模型在实际临床应用中可能面临泛化性能下降的挑战,未来研究需在真实临床拍摄数据上进行验证,以增强模型的鲁棒性与临床适用性。其次,本研究的数据规模对于深度学习任务而言相对有限,尽管已实现初步的模型训练与性能比较,但更大的样本量有助于模型学习更丰富、更稳健的图像特征,并可能进一步提升其分类精度与泛化能力,未来可通过多中心合作扩充高质量临床拍摄数据。最后,本数据集中高、中、低笑线的样本数量呈现不均衡分布,这种分布特征与人群中的自然分布趋势相符。研究显示,在自然微笑状态下,中位笑线最为常见,高位与低位笑线比例相对较低,且其分布可能受到年龄、性别等因素的影响^[18-19]。因此,本研究数据集中的类别不平衡在一定程度上反映了真实人群的固有特点,这为后续开发面向真实人群的智能评估工具提供了具有现实意义的训练基础^[20]。在后续研究中需进一步扩展数据来源,采集规模更大、质量更优的图像数据,以更充分地验证模型的有效性。此外,后续研究可围绕更多类型的微笑任务展开深入探究,以丰富模型的应用场景,从而为口腔美学评估与临床诊断提供更为全面、精准的技术支持。

利益冲突 所有作者均声明不存在利益冲突

作者贡献声明 孟笑菲、龚卓弘:实验设计、分析、解释数据、文章撰写;万耀文、李沛达:实施研究、采集数据、验证实验结果;胡琪琪:数据统计分析、作图;邱龙诗语、刘恒毅:翻译资料、文献汇总、对文章的知识性内容作批评性审阅;谢伟丽:研究指导、论文审阅、支持性贡献

参 考 文 献

- [1] Lukez A, Pavlic A, Trinajstic Zrinski M, et al. The unique contribution of elements of smile aesthetics to psychosocial well-being [J]. *J Oral Rehabil*, 2015, 42(4):275-281. DOI:10.1111/joor.12250.
- [2] Pham TAV, Nguyen PA. Morphological features of smile attractiveness and related factors influence perception and gingival aesthetic parameters [J]. *Int Dent J*, 2022, 72(1):67-75. DOI:10.1016/j.identj.2021.02.001.
- [3] Wang C, Hu WJ, Liang LZ, et al. Esthetics and smile-related characteristics assessed by laypersons [J]. *J Esthet Restor Dent*, 2018, 30(2):136-145. DOI:10.1111/jerd.12356.
- [4] 许砚耕,张艳玲,胡文杰,等.以微笑美观为导向的口腔软组织美学评价方法概述[J]. *口腔医学*, 2025, 45(1):18-24. DOI:

- 10.13591/j.cnki.kqyx.2025.01.004.
- [5] Lee S, Jin G, Park JH, et al. Evaluation metric of smile classification by peri-oral tissue segmentation for the automation of digital smile design [J]. *J Dent*, 2024, 145: 104871. DOI: 10.1016/j.jdent.2024.104871.
- [6] Liu MQ, Xu ZN, Mao WY, et al. Deep learning-based evaluation of the relationship between mandibular third molar and mandibular canal on CBCT [J]. *Clin Oral Investig*, 2022, 26(1): 981-991. DOI: 10.1007/s00784-021-04082-5.
- [7] Zeng P, Song R, Lin Y, et al. Abnormal maxillary sinus diagnosing on CBCT images via object detection and 'straight-forward' classification deep learning strategy [J]. *J Oral Rehabil*, 2023, 50(12): 1465-1480. DOI: 10.1111/joor.13585.
- [8] Cui Z, Fang Y, Mei L, et al. A fully automatic AI system for tooth and alveolar bone segmentation from cone-beam CT images [J]. *Nat Commun*, 2022, 13(1): 2096. DOI: 10.1038/s41467-022-29637-2.
- [9] Gong Z, Li X, Shi M, et al. Measuring the binary thickness of buccal bone of anterior maxilla in low-resolution cone-beam computed tomography via a bilinear convolutional neural network [J]. *Quant Imaging Med Surg*, 2023, 13(12): 8053-8066. DOI: 10.21037/qims-23-744.
- [10] Chen X, Zhou C, Zhu Y, et al. Detecting glaucoma in highly myopic eyes from fundus photographs using deep convolutional neural networks [J]. *Clin Exp Ophthalmol*, 2025, 53(5): 502-515. DOI: 10.1111/ceo.14498.
- [11] Biswas SS. Role of chat GPT in public health [J]. *Ann Biomed Eng*, 2023, 51(5): 868-869. DOI: 10.1007/s10439-023-03172-7.
- [12] Russell BC, Torralba A, Murphy KP, et al. LabelMe: A database and web-based tool for image annotation [J]. *Int J Comput Vision*, 2008, 77(1): 157-173. DOI: 10.1007/s11263-007-0090-8.
- [13] Tjan AH, Miller GD, The JG. Some esthetic factors in a smile [J]. *J Prosthet Dent*, 1984, 51(1): 24-28. DOI: 10.1016/s0022-3913(84)80097-9.
- [14] Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization [C]//*Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*. Venice, 2017: 22-29. <https://ieeexplore.ieee.org/document/8237336>.
- [15] Das N, Hussain E, Mahanta LB. Automated classification of cells into multiple classes in epithelial tissue of oral squamous cell carcinoma using transfer learning and convolutional neural network [J]. *Neural Netw*, 2020, 128: 47-60. DOI: 10.1016/j.neunet.2020.05.003.
- [16] Setzer FC, Shi KJ, Zhang Z, et al. Artificial intelligence for the computer-aided detection of periapical lesions in cone-beam computed tomographic images [J]. *J Endod*, 2020, 46(7): 987-993. DOI: 10.1016/j.joen.2020.03.025.
- [17] Xu X, Liu C, Zheng Y. 3D tooth segmentation and labeling using deep convolutional neural networks [J]. *IEEE Trans Vis Comput Graph*, 2019, 25(7): 2336-2348. DOI: 10.1109/tvcg.2018.2839685.
- [18] Khan M, Kazmi SMR, Khan FR, et al. Analysis of different characteristics of smile [J]. *BDJ Open*, 2020, 6: 6. DOI: 10.1038/s41405-020-0032-x.
- [19] Cunha J, Fernandes GVO, Fernandes JCH, et al. The interference of age and gender on smile characterization analyzed on six parameters: A clinical-photographic pilot study [J]. *Medicina (Kaunas)*, 2023, 59(3): 595. DOI: 10.3390/medicina59030595.
- [20] Shi M, Gong Z, Zeng P, et al. Multi-quantifying maxillofacial traits via a demographic parity-based AI model [J]. *BME Front*, 2024, 5: 0054. DOI: 10.34133/bmef.0054.

(收稿日期: 2025-12-09)

(本文编辑: 王嫒)