

M2T-Net:基于多源数据的跨任务迁移学习舌象诊断方法

曾丽莉, 夏佳楠, 李韶雯, 敬迈科, 赵慧辉, 周雪忠

引用本文

曾丽莉, 夏佳楠, 李韶雯, 敬迈科, 赵慧辉, 周雪忠. [M2T-Net:基于多源数据的跨任务迁移学习舌象诊断方法](#)[J]. 计算机科学, 2025, 52(9): 47-53.

ZENG Lili, XIA Jianan, LI Shaowen, JING Maikē, ZHAO Huihui, ZHOU Xuezhong. [M2T-Net:Cross-task Transfer Learning Tongue Diagnosis Method Based on Multi-source Data](#) [J]. Computer Science, 2025, 52(9): 47-53.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于雷达和视觉融合的多模态空中手写体识别](#)

Multimodal Air-writing Gesture Recognition Based on Radar-Vision Fusion

计算机科学, 2025, 52(9): 259-268. <https://doi.org/10.11896/jsjcx.240400143>

[融合显著视口提取与跨层注意力的全景图像质量评价方法](#)

Panoramic Image Quality Assessment Method Integrating Salient Viewport Extraction and Cross-layer Attention

计算机科学, 2025, 52(9): 249-258. <https://doi.org/10.11896/jsjcx.241000108>

[数据分类分级技术研究综述](#)

Survey of Data Classification and Grading Studies

计算机科学, 2025, 52(9): 195-211. <https://doi.org/10.11896/jsjcx.240800149>

[基于局部增强傅里叶神经算子的偏微分方程求解方法](#)

Partial Differential Equation Solving Method Based on Locally Enhanced Fourier Neural Operators

计算机科学, 2025, 52(9): 144-151. <https://doi.org/10.11896/jsjcx.240700122>

[基于药物子结构与蛋白质三维图信息的化合物-蛋白质相互作用预测](#)

Graph-based Compound-Protein Interaction Prediction with Drug Substructures and Protein 3D Information

计算机科学, 2025, 52(9): 71-79. <https://doi.org/10.11896/jsjcx.250100116>

M2T-Net:基于多源数据的跨任务迁移学习舌象诊断方法

曾丽莉¹ 夏佳楠¹ 李韶雯¹ 敬迈科¹ 赵慧辉² 周雪忠¹

1 北京交通大学计算机科学与技术学院 北京 100044

2 北京中医药大学中医学院 北京 100029

(22120463@bjtu.edu.cn)

摘要 冠心病是临床常见的心血管疾病,冠脉介入术是其常见治疗方法之一。然而,糖尿病是冠心病的危险因素,与冠心病合并会显著增加治疗风险,尽早诊断和采取相应措施对这类患者具有重要的临床意义。临床指标是目前诊疗冠心病及其合并病的重要参考依据,而这些指标的获取大多是有创的。舌象作为人体健康的外在表现,不仅反映舌色、苔色等特征,还与心脏的各种生理和病理特征关联。深度学习的发展为客观化与可重复性获取舌象表征提供了帮助。然而,现有舌象分类方法受限于数据集标签的单一性,导致模型泛化能力不足。为此,提出了一种基于多源数据的跨任务迁移学习舌象诊断方法 M2T-Net。该方法包括两个阶段:在多源数据的预训练阶段,获取不同任务下的高质图像编码器;在跨任务迁移阶段,结合交叉注意力机制,融合不同任务的特征表示,用于疾病分类。实验表明,M2T-Net 模型在冠心病和冠心病伴随糖尿病两种人群的分类任务上的分类准确率达到 93%,优于现有先进方法,具备较强的泛化能力与实用性,并且跨任务获得疾病表征更符合中医舌诊的整体观诊断思想,为舌象分析领域提供了更具实用性的解决方案。

关键词: 迁移学习;舌象诊断;交叉注意力;深度学习;舌诊智能化

中图分类号 TP391

M2T-Net:Cross-task Transfer Learning Tongue Diagnosis Method Based on Multi-source Data

ZENG Lili¹, XIA Jianan¹, LI Shaowen¹, JING Maikē¹, ZHAO Huihui² and ZHOU Xuezhong¹

1 School of Computer Science and Technology, Beijing Jiaotong University, Beijing 100044, China

2 School of Traditional Chinese Medicine, Beijing University of Chinese Medicine, Beijing 100029, China

Abstract Coronary artery disease is a common clinical cardiovascular disease, and coronary intervention is one of its common treatment methods. However, diabetes mellitus is a risk factor for coronary artery disease, and the combination of diabetes mellitus and coronary artery disease significantly increases the risk of treatment, so early diagnosis and corresponding measures are of great clinical significance for these patients. Clinical indicators are important references for the diagnosis and treatment of coronary heart disease and their comorbidities, and most of these indicators are invasive. Tongue image, as an external manifestation of human health, not only reflects the tongue color, moss color and other characteristics, but also correlates with various physiological and pathological features of the heart. The development of deep learning provides assistance for objectivized and reproducible acquisition of tongue representations. However, existing tongue image classification methods are limited by the singularity of dataset labels, which leads to the lack of model generalization ability. To this end, a cross-task migration learning tongue diagnosis method M2T-Net based on multi-source data is proposed. Specifically it consists of two phases, in the pre-training phase of multi-source data, high quality image encoders under different tasks are acquired. In the cross-task migration phase, the feature representations from different tasks are fused for disease classification by combining the cross-attention mechanism. Experiments show that the performance of the M2T-Net model in the classification tasks of coronary heart disease and coronary heart disease accompanied by diabetes mellitus reaches a classification accuracy of 93%, which is better than the existing state-of-the-art methods, with strong generalization ability and practicality, and the cross-task acquisition of disease representations is more in line with the holistic diagnostic idea of Chinese medicine tongue diagnosis, which provides a more practical solution for the field of tongue image analysis.

Keywords Transfer learning, Tongue image diagnosis, Cross attention, Deep learning, Intelligent tongue diagnosis

到稿日期:2024-10-11 返修日期:2025-02-08

基金项目:中央高校基本科研业务费(2024JBMC007);国家重点研发计划(2023YFC3502604)

This work was supported by the Fundamental Research Funds for the Central Universities(2024JBMC007) and National Key Research and Development Program of China(2023YFC3502604).

通信作者:夏佳楠(xiajn@bjtu.edu.cn)

1 引言

冠心病是一种慢性疾病,涉及糖尿病、高血压和年龄等不同种类的危险因素的组合。它是动脉粥样硬化使得冠状动脉变窄或者阻塞,导致血液供应减少和心肌缺氧^[1]。研究表明^[2],冠心病是一项全球公共卫生挑战,在身心健康、生活质量、工作效率和医疗保健成本方面给个人、家庭和社会带来了沉重的负担。同时,糖尿病是发生冠心病以及已确诊冠心病患者心血管疾病的一个非常普遍的危险因素,伴随糖尿病患者的手术难度与风险随之增加。利用机器学习技术实现冠心病和冠心病伴随糖尿病两种人群的精准分类,将为患者制定个性化治疗方案提供依据,确保患者得到最佳治疗^[3]。中医以辨证用药为特色,不同证型采用不同治疗方剂,从四诊(望、闻、问、切)中收集信息,经过综合分析后,判断患者的证候。

舌诊作为中医疾病诊断的重要手段,以其便捷性和显著的诊断价值在临床上得到广泛应用。舌象能够反映多种信息,主要包括颜色、纹理和形状等特征,在舌体上表现为舌色和舌形,而在舌苔上则体现为苔色和苔质^[4]。然而,传统舌诊容易受到医生主观经验的影响,并缺乏客观评判标准,这限制了其发展。为缓解主观因素对诊断结果的影响,研究者开始将舌诊与深度学习技术相结合,旨在实现舌头表征的客观化与智能化,以辅助医生提高诊断效率。

具体地,舌头表征获取任务主要分成舌部区域分割和舌头特征分类^[5]。舌头特征分类是为了识别舌象类别从而辅助医生进行下一步的诊断。然而,舌部特征的类间差异集中于舌色、舌质、苔色、苔质等细微差别,因此舌部特征分类是一项艰巨的细粒度分类任务^[6]。不同类型的舌像提示冠心病患者处于不同的发展阶段和严重程度^[7],不同病情的患者如果采用相同的治疗方案,势必会影响部分患者的治疗效果。获取患者准确的舌像分类结果可以保证患者得到与病情对应的治疗。

在临床上,冠心病患者的舌象根据分型有所不同,为了提升深度学习模型的能力,从数据的角度出发,需要构建高质量且多标签的数据集以满足复杂任务的需求。在中医四诊中,舌象作为复杂的多标签数据,不仅反映舌色、舌形和舌苔等基本特征,还可能与多种疾病状态相关联^[8-9]。然而,目前开源的舌部数据集数量较少,且每个数据集的标签往往局限于单一任务,例如舌色分类、舌苔分类或疾病检测等。理想情况下,多标签舌象数据集能够为每个样本提供多个相关的标签信息,例如舌象、舌色对应的疾病,使得模型能够更全面地学习和理解数据的多样性与复杂性。临床上,构建高质量的多标签舌象数据集面临诸多挑战,如数据收集难度大、标签不一致以及数据量不足等。但在实际医学诊断中,仅依赖单一特征的简单分类模型,对医生提供的实质性帮助有限。因此,本研究致力于解决单标签数据集在不同任务间的迁移问题,结合各个单标签分类模型提取特征的能力,并运用到新任务上,实现对复杂舌象的全面理解,以期实现更有效的舌诊辅助工具。

深度学习在舌象分类领域的应用高度依赖样本的质量和数量,而收集大量舌象且带标签的样本较为困难,这些问题导

致舌象分类模型缺乏足够的跨任务泛化能力。如何在多个相关任务上共享知识?迁移学习中的域适应是一种常见的应对策略,它可以辅助基于医学图像的有效数据驱动诊断建模^[10]。迁移学习已被成功应用于各种实际应用中,帮助有限的资源训练深度神经网络^[11-13]。然而,传统的迁移方法直接重用来自公共数据库(如 ImageNet^[14])的知识,可能不适用于特定的医学数据,因为医学图像中的模式具有独特的结构和形态特征。事实证明,如果不考虑源域和目标域样本之间的相似度,迁移学习将导致模型性能低下^[15-16]。利用已有的其他领域知识,有效地转移知识以及转移的是什么,是迁移学习所需解决的重要问题^[17]。对此,Kumar 等^[18]使用医学图像对在一般图像数据集上预训练的 CNN 进行微调,创建了一个新的优化特征提取器,它能够调整自然图像中学到的通用特征,以更好地适应医学成像模式。针对肺癌图像,Wang 等^[19]采用预训练的残差神经网络作为迁移学习模型,在此基础上增加全连接层以增强模型的迁移能力,并在山东省医院收集的肺癌数据集上进行微调,用于肺癌图像的特征提取和分类任务。基于特征的迁移方法提取通用性更强的域不变特征,使得后续分类任务可以不受域变化的影响,能适应域差异较大的迁移场景。尽管现有的迁移学习方法在特定任务上取得了不错的成绩,但它们通常仅针对同一任务不同域进行优化,未能充分利用不同任务之间的潜在关联。这种局限性使得在处理多样化数据集时,模型在泛化能力和性能上仍显不足。因此,考虑舌色分类、舌苔分类、病人诊断任务之间的特殊关联性,以及舌象存在的域不变特征,本文使用基于特征的迁移学习方法来提高舌象分类模型的泛化能力。

综上所述,模型应关注单标签数据特征的提取能力,以便跨任务知识融合顺利进行。同时,将迁移学习的优势整合其中,解决缺乏宝贵的中医数据样本标签的问题,从而降低训练模型的成本。本文的思想是:如果任务之间有相关性,那么通过对对应任务提取出的特征也有相关性,利用融合相关性的特征,能够帮助模型区分不同病症患者,从而提升模型分类准确率。本文提出基于多源数据的跨任务迁移学习舌象诊断方法 M2T-Net,并以冠心病和冠心病伴随糖尿病患者分类为例加以说明。在第一个阶段,利用多个开源单标签舌象数据集独立进行预训练,获取特定任务的图像编码器。第二个阶段,结合多任务编码器进行迁移学习,并引入交叉注意力机制来融合不同编码器的特征,实现更有效的冠心病和冠心病伴随糖尿病两种疾病的分类。本文模型在更好地训练模型的同时,解决了临床舌象标签缺乏的问题。模型不仅能够捕捉到不同数据源之间的共性,还能通过多任务学习的方式增强模型对复杂舌象的适应性,提升疾病分类精度。希望通过这种创新的思路,克服舌象数据标签稀缺的限制,对舌体、舌苔等主要属性同时进行识别分析,并落地于疾病分类。这样训练的模型更符合中医舌诊的整体观诊断思想,也能够为舌象分析领域提供更具实用性的解决方案。

2 相关工作

2.1 舌象诊断方法

舌诊主要包括对舌体和舌苔进行识别,目前已有大量基

于舌象的诊断方法。Hou 等^[20]在 CaffeNet 的结构上进行修改,依据舌色对舌象进行分类。Wang 等^[21]构建了 1548 张不同设备捕获的舌头图像,使用 ResNet34 CNN 架构提取特征并进行分类,模型性能达到 90%。舌象可以反映舌体颜色、苔色、舌苔厚度等多种信息,也就是说,舌头图像是多标签数据。Zhang 等^[22]将多标签学习应用于舌部图像分类,将舌苔和舌体分离后提取颜色特征和纹理特征,并使用多标签学习算法进行分类。Huang 等^[6]提出 TSC-WNet,其包含两个子网络,其中 TSC-Net 是简单有效的分类主干;此外,他们为舌头构建了分割网络 TSC-UNet,用于提供互补的有益特征,以提高网络对舌头大小和形状的分类性能。

尽管前人的研究已取得显著成果,但计算机辅助的舌象诊断在很大程度上仍然是一个值得深入探索的领域。由于医疗数据集的隐私性,现有舌诊文献中可用的数据集数量和任务分类相对有限。因此,本文旨在利用舌象的多标签特性,提出一个跨任务迁移的舌诊分类网络,以最大化融合已开源的舌象数据。

2.2 在医学场景中的迁移学习

迁移学习能够有效克服医学图像标注数据稀缺的问题,提升分类效果,因此成为该领域具有潜力的研究方向^[23]。

目前医学场景中的迁移学习方法主要集中在基于实例的方法和基于模型的方法两大类。基于实例的方法旨在使用特定的权重调整策略,挑选源域中的实例补充到目标域中。DeMoura 等^[24]基于迁移学习使用不同的特征选择策略,从目标 OCT 图像中选择最相关的深度特征子集,实现糖尿病性黄斑水肿患者的分类。基于网络的深度迁移学习主要将经过充

分训练的网络重用到目标域。它将源域神经网络的前端层视为多功能的特征提取器,其结构和参数将成为新网络的一部分,但有时会在再训练中冻结。Khan 等^[25]使用由自然图像组成的大型基准数据集预训练的 VGG,通过分层调整对网络进行微调,利用一组预定义的层在 MRI 图像上进行训练,以预测阿尔兹海默症。Zhang 等^[26]提出了一种基于实例和基于网络相结合的两阶段深度迁移策略,以解决异常舌头检测中数据不足的问题。第一阶段获取不平衡数据集的可用实例,学习公共基本特征并从更相似的领域转移知识,第二阶段转移结构特征。

现有的方法通常侧重于提取域不变特征进行迁移。受此启发,不同任务之间存在潜在关联,如果模型能够考虑这种关联并在任务间进行迁移,将有助于提升模型性能。舌象是一种复杂的多标签数据,如何将舌象诊断、舌色分类、苔色分类等不同任务关联起来,挖掘符合中医诊断逻辑的疾病特征,是本文要解决的核心问题。

3 模型与方法

3.1 整体框架

本文提出了一种基于多源数据融合的跨任务迁移学习舌象诊断方法——M2T-Net,其框架如图 1 所示。该方法主要分为两个核心部分:1)多源数据预训练阶段,通过不同任务标签的数据集构建多个图像编码器;2)跨任务迁移阶段,利用预训练的图像编码器提取特定的图像特征,并通过交叉注意力机制对齐并融合不同任务之间的舌象表征。最终,在下游任务中验证其在冠心病及冠心病伴随糖尿病诊断中的有效性。

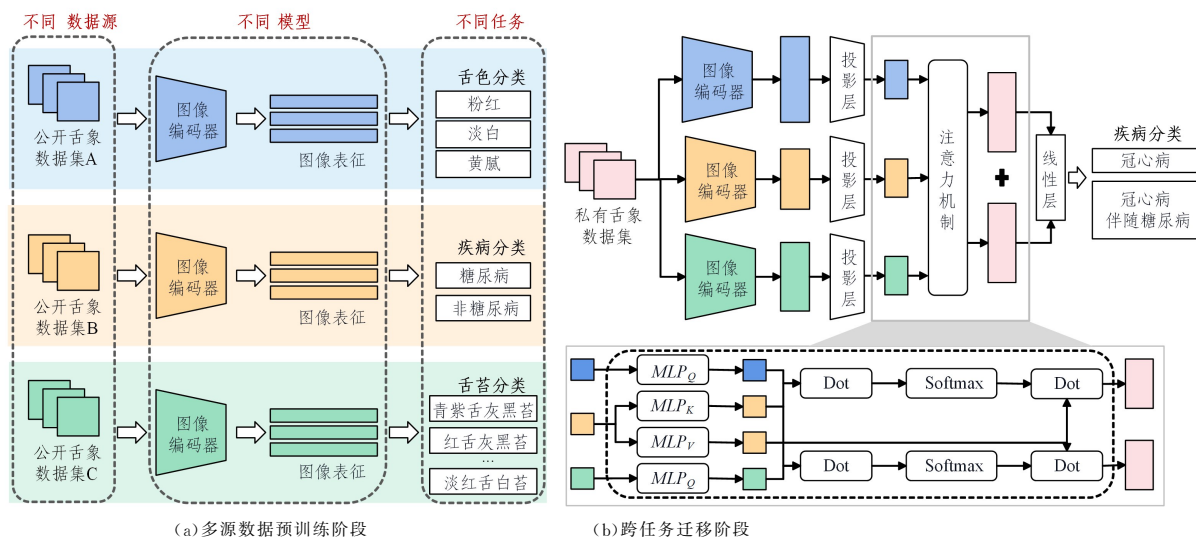


图 1 本文模型总体框架图

Fig. 1 Overall framework of the proposed model

3.2 多源数据预训练阶段

在多源数据预训练阶段,由于开源舌象数据集相对匮乏且多集中于单一标签,因此将现有的开源数据集问题划分为三大类。

1)舌色分类任务。该任务包括 3 种主要标签:粉红、淡白色和黄腻色。大多数开源舌色分类数据集均涵盖这 3 类标签,少数数据集增加了如绛紫色等其他细分标签。为了增强

模型的可迁移性,选择舍弃这些少数情况,专注于上述三大类标签。在飞浆平台上收集到的舌色分类数据集共有 2262 张图片,经过旋转、裁切等数据增强操作,在 ResNet, ViT, CNN 等经典网络结构上进行分类,最终选定了分类性能达到 96% 的 ViT 作为舌色特征提取器。

2)疾病分类任务。在中医四诊中,糖尿病患者的舌象与非糖尿病患者之间存在显著差异。健康人群的舌色通常表现

为淡红色,而糖尿病患者身体出现病理变化时,其舌象会相应发生改变。例如,舌苔变厚可能提示血糖控制不佳,而早期糖尿病患者的舌头往往呈偏红色,表明体内热量过高。因此,有多个开源数据集专门收集糖尿病患者的舌象,以便与健康人群或其他疾病患者进行区分。在 Kaggle 平台收集了 1100 张舌部患者的图片,与舌色分类任务类似,采用 ResNet, ViT, CNN 等经典网络进行分类,最终选择了分类性能达到 99% 的 ResNet101 模型作为糖尿病预测的特征提取器。

3) 舌苔分类任务。舌部区域的特征不仅包括舌头的颜色,还涉及苔色和苔质等特征。对于每张舌象,构建了一个多标签数据集,包含 4 个标签:舌色、苔色、苔质 1 和苔质 2。舌色标签包括“青紫舌”“绛舌”“红舌”“淡红舌”和“淡白舌”5 种;苔色则定义为“灰黑苔”“黄苔”和“白苔”3 种情况;苔质 1 描述苔的薄厚程度,包括“厚苔”“薄苔”和“正常”3 种状态;苔质 2 则补充是否为“腻苔”的特殊情况。构建多标签数据集后,针对不同任务进行多分类训练,与舌色分类和疾病分类不同,舌苔分类任务选定了 4 个多分类头,以提高模型对舌色、苔色和苔质的识别能力。本文选择了分类准确率为 84.25% 的 EfficientNet 网络,以便后续对舌苔特征的获取。

这 3 个预训练任务从多源数据角度出发,旨在最大化现有资源对舌部图像的处理。尽管总体任务均为分类,但从临床实践的角度来看,各任务之间仍存在明显差异。本文将 3 个任务的标签分离开来,利用 3 个独立网络进行训练,获得了 3 个各具侧重的图像编码器,以提取特定任务更为关注的图像特征。

3.3 跨任务迁移阶段

在准备好基于多源舌部图片预训练的图像编码器后,进入第二阶段,也就是跨任务迁移阶段。在该阶段中,主要考虑利用 3 个不同任务的图像编码器在私有数据集上进行迁移,同时利用交叉注意力机制对齐不同任务之间的表征。

具体而言,舌象 i 先分别通过舌色分类的图像编码器、疾病分类的图像编码器,以及舌苔分类的图像编码器,提取对应的特征 Col_i, Dis_i, Coa_i 。值得注意的是,每个图像编码器均保留最后一层隐藏层的输出作为其图像特征表征,并且通过 3 种编码器获得的特征在特征空间上分布不同。为了对齐不同任务训练获得的特征,利用 3 个投影层对特征形状进行调整, Dis_i 通过线性层生成键 D_k 和值 D_v , Col_i 通过线性层生成查询 C_q 。将 D_k 与 C_q 进行点乘,通过点乘衡量舌色特征 Col_i 对于疾病特征 Dis_i 的相关性,再经过 softmax 函数归一化生成注意力权重,之后将 D_v 和注意力权重相乘,最终得到与舌色特征最相关的疾病特征 f_1 ,具体如式(1)所示:

$$f_1 = \text{softmax}\left(\frac{C_q \cdot D_k^T}{\sqrt{d_k}}\right) \cdot D_v \quad (1)$$

其中, d_k 是特征 D_k 的第一个维度,即隐藏层的长度。同理,获得与舌苔特征最相关的疾病特征 f_2 。最后将两个特征进行连接,获得舌色与舌苔关注的疾病特征,再将其映射到二维概率空间中,得到病人是患有冠心病或者冠心病伴随糖尿病的结果。模型训练过程使用交叉熵损失进行约束。整个过程不仅有效实现了跨任务特征的对齐,也为后续的临床诊断提供了坚实的数据基础。

4 实验研究与分析

4.1 舌象数据集

本文基于飞浆平台和 Kaggle 平台收集了开源的舌象图片,并根据任务类型将数据集分为 3 类:舌色分类、舌苔分类和疾病分类。具体而言,舌色分类数据集包含 4409 张图像,舌苔分类数据集包含 2088 张图像,而疾病分类数据集则包含 1100 张图像。每种标签的舌象数量如表 1 所列。需要指出的是,舌色分类和疾病分类均为单标签分类任务,而舌苔分类为多标签分类,4 类标签的可能取值以“/”分隔。

表 1 开源实验数据情况

标签	数目	任务
粉红	1967	舌色分类
淡白	1662	舌色分类
黄腻	780	舌色分类
糖尿病/非糖尿病	550/550	疾病分类
淡白舌/淡红舌/ 红舌/青紫色/绛舌	118/1314/ 506/70/80	舌苔分类
黄苔/白苔/灰黑苔	531/1466/91	舌苔分类
薄苔/厚苔/其他	765/1136/187	舌苔分类
腻苔/非腻苔/其他	1363/538/187	舌苔分类

为了进一步验证所提出的基于多源数据融合的跨任务迁移学习舌象方法的有效性,本文从舌象库中选取了 675 幅舌部图像。这些图像由专业设备采集,并且附有对应的舌色和舌苔描述。数据集的详细信息如表 2 所列。在实验中,选择疾病分类任务进行验证,表 2 中列出的其他标签是从数据集中相应的文本描述中提取而来。鉴于部分数据的标签情况较为复杂,本文仅列出可清晰判定的几种标签及其对应的图像数量。

表 2 私有数据集信息

标签	数目	标签	数目
冠心病	422	白苔	120
冠心病伴糖尿病	253	淡红	412
薄苔	317	黄腻	10
厚苔	245	淡白	14
黄苔	205	绛舌	96

在实验过程中,每个数据集中的舌象图像均要求舌体充分暴露且舌面平展。考虑到数据的有限性,本文将数据集按照每类标签 8:2 的比例进行训练集和测试集的划分,并进行五折交叉实验,以验证所提方法的可靠性。

4.2 实验设置

实验环境配置为:本实验在 LINUX64 Ubuntu 20.04.3 LTS 操作系统上进行,硬件配置为 Intel^(R) Xeon^(R) Silver 4210 CPU 和 NVIDIA GeForce RTX 2080 Ti 显卡。实验框架为 PyTorch。

4.3 分类评价准则

本文基于样本进行评价,对于分类模型的有效性选择准确率(Accuracy)、精确率(Precision)和召回率(Recall)作为评价指标。

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

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

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

其中, TP 表示实际为正且被预测为正的样本数, TN 表示实际为负且被预测为负的样本数, FP 表示实际为负但被预测为正的样本数, FN 表示实际为正但被预测为负的样本数。

4.4 实验结果分析

4.4.1 对比实验

本研究旨在评估基于多源数据的跨任务迁移学习舌象诊断方法 M2T-Net 的有效性。模型的创新点在于,利用多源单标签舌象数据集进行预训练,并借助预训练的特征提取器在私有数据集上针对冠心病(CHD)和冠心病伴随糖尿病(CHD+DM)患者开展分类实验。模型通过跨任务迁移学习,挖掘舌象任务之间的关联性,进一步提升了疾病识别的准确性。

为了验证模型的必要性和有效性,选取多种经典的图像分类模型作为对比,包括卷积神经网络(CNN)和视觉 Transformer(ViT)等,并对其在冠心病(CHD)和冠心病伴随糖尿病(CHD+DM)分类任务中的性能进行了全面分析,具体结果如表 3 所列。为确保实验的公平性,所有模型均在相同的私有数据集上进行训练和测试。值得注意的是,对比模型均加载了已开源的预训练的权重,并在私有数据集上通过迁移学习进行微调操作,最终保留各模型在私有数据集上的最佳结果。加载预训练权重并在特定任务的数据集上进行微调的实验设计,符合迁移学习的理论依据,同时保证了对比模型的训练流程与 M2T-Net 的训练流程一致,能够更加真实地反映各模型在处理类似任务时的性能差异,为验证 M2T-Net 的有效性提供了有力支持。

由于深度学习模型的性能依赖于标签质量,而本文所使用的私有数据集标签仅涵盖冠心病和冠心病伴随糖尿病两类,因此未在对比模型中引入来源于开源数据集的跨任务标签,以避免对模型分类性能产生干扰。为减小过拟合的风险,采用交叉验证的方法,并在训练过程中监控每个模型的分精度、损失值以及最终的分性能指标。

与经典卷积网络 VGG16 和 AlexNet 相比,VGG16 的准确率为 0.85,精确率为 0.89,召回率为 0.87,整体表现较为逊色。与 M2T-Net 相比,VGG16 在 3 项指标上均明显落后,特别是召回率(差异 8%),表明其识别冠心病伴随糖尿病的能力较弱。AlexNet 在召回率上表现出色,达到了 0.95,优于 M2T-Net(0.93),但其精确率和准确率分别为 0.90 和 0.89,与 M2T-Net 相比有所差距。虽然 AlexNet 能够有效识别冠心病伴随糖尿病的人群,但整体分类能力仍落后于 M2T-Net。原因可能在于,其仅适用于单一的较为浅层的特征提取结构,限制了模型对复杂疾病特征的捕捉。而 M2T-Net 结合不同特征提取器,对于冠心病伴随糖尿病与冠心病人群中细微的舌象差异更为敏感,性能更好。

ResNet 巧妙的残差连接设计能有效应对梯度消失问题。ResNet18 的整体表现与 M2T-Net 较为接近,准确率和召回率均达 90%以上,但精确率稍低。ResNet50 尽管在精确率上

与 M2T-Net 持平,但准确率和召回均低于本文模型。更深的 ResNet101 并未带来预期的性能提升,这表明模型深度过大可能导致过拟合,而无法充分捕捉疾病特征。

此外,实验对比了目前主流的网络 ViT 和 ConvNeXt_Tiny。ViT 通过自注意力机制捕捉图像全局特征,尽管其召回率达到了 0.93,与 M2T-Net 相当,但其准确率和精确率显著低于 M2T-Net。这说明,ViT 在较复杂的医学图像分类任务中未能充分发挥优势。而 ConvNeXt_Tiny 作为一种现代卷积神经网络,其表现接近 M2T-Net,但准确率略低于 M2T-Net,说明 M2T-Net 在融合多模态数据时更具优势。

与 DenseNet 对比,该网络使用密集连接层实现高效的特征传递,避免了信息的丢失。DenseNet 在精确率上表现突出,达到了 0.95。然而,DenseNet 的准确率和召回率均比 M2T-Net 低 3%以上,这是由于 DenseNet 对于冠心病伴随糖尿病的识别能力较弱。实验表明,尽管 DenseNet 在局部特征提取方面表现出色,但 M2T-Net 在整体特征融合和分类任务上表现得更为全面和鲁棒。

综上所述,M2T-Net 在准确率和召回率上表现最佳,表明该模型能够较为精准地捕捉大多数真实的冠心病伴随糖尿病患者。尽管其精确率略低于 DenseNet,但在实际应用中,这种差异并不会显著影响其临床价值。相反,对于误召回的冠心病患者,这一结果提示其可能具有较高的潜在糖尿病风险,从而为早期疾病筛查和预警提供了重要参考。

表 3 对比实验结果

方法	准确率	精确率	召回率
EfficientNet	0.90	0.89	0.92
AlexNet	0.89	0.90	0.95
VGG16	0.85	0.89	0.87
Vit	0.82	0.81	0.93
ResNet18	0.90	0.92	0.93
ResNet50	0.88	0.92	0.90
ResNet101	0.85	0.90	0.88
ConvNeXt_Tiny	0.90	0.92	0.93
DenseNet	0.90	0.95	0.89
M2T-Net	0.93	0.93	0.95

4.4.2 消融实验

为了进一步观察多源跨任务的表现,设置了消融实验,主要针对以下两个关键问题。

1) 迁移学习对模型性能的贡献。为了验证利用多源跨任务迁移的效果,从所提方法中移除与相关任务的开源舌象数据集,用已开源预训练权重代替。

2) 注意力机制对模型性能的贡献。为了验证跨任务迁移阶段中是否有必要利用注意力机制对不同任务特征融合,将不同特征编码器提取的特征直接拼接,并设置为“未加注意力”。

消融实验的结果如表 4 所列,交叉注意力机制和迁移学习对 M2T-Net 的性能提升起到了关键作用。首先,在移除交叉注意力机制的情况下,模型的准确率下降至 0.86,精确率下降至 0.88,而召回率为 0.91。尽管召回率有所保持,但模型整体性能显著下降,说明交叉注意力在对齐不同任务特征方面具有重要贡献。

表 4 消融实验结果

Table 4 Ablation results

方法	准确率	精确率	召回率
未加注意力	0.86	0.88	0.91
未加迁移学习	0.88	0.96	0.86
M2T-Net	0.93	0.93	0.95

其次,在没有应用迁移学习的情况下,模型的精确率达到了0.96,高于M2T-Net的0.93,但也伴随着召回率的大幅下降,仅为0.86。这表明,尽管迁移学习对精确率的提升有限,但其对提升模型的整体平衡表现(即在精确率和召回率之间的权衡)起到了至关重要的作用。

引入迁移学习后,模型识别冠心病伴随糖尿病患者的能力得到了显著提升,从而有效减少了漏检的风险。尽管迁移学习的应用导致了精确率的轻微下降,但召回率却显著提高。精确率的下降主要是由于模型将部分仅患有冠心病的患者误判为冠心病伴随糖尿病患者,表明这些患者可能存在潜在的健康风险,因此具有一定的预警意义。精确率与召回率之间的权衡反映了迁移学习在优化模型整体性能方面的价值,特别是在临床实践中,能够有效降低漏检的可能性,这对于冠心病患者的健康管理具有重要的意义。通过这种平衡,迁移学习不仅提升了诊断的全面性,还增强了冠心病伴随糖尿病的早期预警的能力,有助于及早发现潜在危险患者,进而改善医疗决策和患者预后。

总体来看,M2T-Net通过交叉注意力机制融合多源域数据特征,实现了更为均衡的性能表现,准确率、精确率和召回率分别达到了0.93,0.93和0.95,明显优于消融后的模型。这表明,在不同任务的场景下的迁移学习,使用交叉注意力对齐特征之间差异,能够有效提升模型在疾病分类任务中的综合表现。

结束语 本文基于舌象的多标签特性,针对舌象分类、舌苔分类以及疾病分类等多任务学习问题,提出了一种新的多源数据的跨任务迁移学习舌象诊断方法——M2T-Net。模型在第一阶段通过对多个开源舌象数据集进行预训练,分别获得针对特定任务的图像编码器。第二阶段,采用交叉注意力机制融合不同任务的特征,从而提升模型的多任务学习能力。最终,M2T-Net成功应用于冠心病和冠心病伴随糖尿病患者的分类任务,实验结果表明,模型在该任务中的准确率达到93%,相比现有的其他先进网络表现更为优异,同时有效平衡了精确率和召回率。

尽管如此,本文方法仍存在一些局限性。由于医疗数据涉及隐私保护,开源舌象数据的数量相较于自然图像数据依然较少。虽然本研究采用了五折交叉验证来提高模型的可靠性,但数据量的不足仍然是限制模型性能提升的一个因素。为了进一步增强M2T-Net的应用效果,未来考虑结合多种数据增强技术,扩展数据集规模,提升数据质量,从而进一步发挥M2T-Net的优势,推动舌象分析技术在实际医疗诊断中的应用。

参考文献

[1] HEUSCH G. Myocardial ischaemia-reperfusion injury and car-

dioprotection in perspective [J]. *Nature Reviews Cardiology*, 2020,17(12):773-789.

- [2] GUO J,WANG H,LI Y,et al. Nanotechnology in coronary heart disease [J]. *Acta Biomaterialia*,2023,171:37-67.
- [3] ZHOU H,LI L,ZHAO H,et al. A large-scale,multi-center urine biomarkers identification of coronary heart disease in TCM syndrome differentiation [J]. *Journal of Proteome Research*, 2019,18(5):1994-2003.
- [4] ZHANG D,ZHANG H,ZHANG B. Introduction to tongue image analysis[M]// *Tongue Image Analysis*. Singapore:Springer, 2017:3-18.
- [5] XU Q,ZENG Y,TANG W,et al. Multi-task joint learning model for segmenting and classifying tongue images using a deep neural network [J]. *IEEE Journal of Biomedical and Health Informatics*,2020,24(9):2481-2489.
- [6] HUANG Y,LI X,ZHENG S,et al. Tongue size and shape classification fusing segmentation features for traditional Chinese medicine diagnosis [J]. *Neural Computing and Applications*, 2023,35(10):7581-7594.
- [7] WANG J,JIANG L,CHANG T,et al. Advances in the Application of Traditional Chinese Medicine Tongue Diagnosis in Coronary Heart Disease [J]. *Practical Journal of Cardiology*,2024,32(3):99-102.
- [8] XU J. Color atlas of tongue diagnosis in traditional Chinese medicine [M]. Shanghai:Shanghai University of TCM Press,2009.
- [9] LI Y,LUO Y,ZHU Z. Tongue Image Feature Analysis Based on Deep Learning [J]. *Computer Science*,2020,47(11):148-158.
- [10] YOSINSKI J,CLUNE J,BENGIO Y,et al. How transferable are features in deep neural networks? [EB/OL]. https://proceedings.neurips.cc/paper_files/paper/2014/file/532a2f85b6977104bc93f8580abb330-Paper.pdf.
- [11] YIN X,YU X,SOHN K,et al. Feature transfer learning for face recognition with under-represented data [C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.
- [12] WEI W,MENG D,ZHAO Q,et al. Semi-supervised transfer learning for image rain removal[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.
- [13] PERERA P,PATEL V M. Deep transfer learning for multiple class novelty detection[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.
- [14] DENG J,DONG W,SOCHER R,et al. Imagenet: A large-scale hierarchical image database[C]// 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009.
- [15] XIE M,JEAN N,BURKE M,et al. Transfer learning from deep features for remote sensing and poverty mapping[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2016.
- [16] XU R,LI G,YANG J,et al. Larger norm more transferable: An adaptive feature norm approach for unsupervised domain adaptation[C]// 2019 IEEE/CVF International Conference on Computer Vision. 2019.
- [17] CSURKA G. A comprehensive survey on domain adaptation for visual applications [J]. *arXiv:1702.05374*,2017.
- [18] KUMAR A,LYNDON D,KIM J,et al. Subfigure and Multi-Label Classification using a Fine-Tuned Convolutional Neural Net-

- work[C]//Proceedings of the CLEF. 2016.
- [19] WANG S,DONG L,WANG X,et al. Classification of pathological types of lung cancer from CT images by deep residual neural networks with transfer learning strategy [J]. *Open Medicine*, 2020,15(1):190-197.
- [20] HOU J,SU H Y,YAN B,et al. Classification of tongue color based on CNN[C]//Proceedings of the 2017 IEEE 2nd International Conference on Big Data Analysis(ICBDA). 2017.
- [21] WANG X,LIU J,WU C,et al. Artificial intelligence in tongue diagnosis: Using deep convolutional neural network for recognizing unhealthy tongue with tooth-mark [J]. *Computational and Structural Biotechnology Journal*,2020,18:973-980.
- [22] ZHANG X,ZHANG J,HU G,et al. Preliminary study of tongue image classification based on multi-label learning[C]//Proceedings of the Advanced Intelligent Computing Theories and Applications;11th International Conference. 2015.
- [23] YU Y. Advances in application of artificial intelligence in medical image analysis [J]. *Chinese Journal of Medical Imaging Technology*,2019,35(12):1808-1812.
- [24] DE MOURA J,NOVO J,ORTEGA M. Deep feature analysis in a transfer learning-based approach for the automatic identification of diabetic macular edema[C]//Proceedings of the 2019 International Joint Conference on Neural Networks. 2019.
- [25] KHAN N M,ABRAHAM N,HON M. Transfer learning with intelligent training data selection for prediction of Alzheimer's disease [J]. *IEEE Access*,2019,7:72726-72735.
- [26] ZHANG X,CHEN Z,GAO J,et al. A two-stage deep transfer learning model and its application for medical image processing in Traditional Chinese Medicine [J]. *Knowledge-based Systems*, 2022,239:108060.



ZENG Lili, born in 2000, postgraduate, is a member of CCF(No. R0493G). Her main research interests include medical image analysis and domain generalization.



XIA Jianan, born in 1990, Ph.D, lecturer, master's supervisor, is a member of CCF(No. P2378M). Her main research interests include time series analysis and medical image analysis.

(责任编辑:何杨)