

深度学习舌诊模型在消化道疾病中的应用研究进展

林映清¹, 刘熙荣^{2*}

¹广西中医药大学第一临床医学院, 广西 南宁

²广西中医药大学第一附属医院消化内镜诊疗区, 广西 南宁

收稿日期: 2026年4月19日; 录用日期: 2026年5月12日; 发布日期: 2026年5月21日

摘要

舌诊作为中医四诊中望诊的代表手段之一, 可令医者做到司外揣内, 然而传统舌诊常受限于医生的经验和主观性而无法充分发挥其作用, 基于深度学习构建的各类舌诊模型的出现为上消化道疾病的早期筛查和辨证施治提供了智能化支持。本文探究舌诊客观化发展历程, 检索了近10年来发表的高质量研究, 对舌象采集及处理、模型网络架构、模型优化等方面的进展及深度学习舌诊模型在消化道疾病中的应用研究进行了系统综述, 发现现有研究多仍存在如样本量受限、多模态模型较少等不足, 未来需朝着多中心、大样本、标准化采集迈进, 可将舌诊模型与胃镜图像结合综合进行预测, 提高智能舌诊模型的预测精度, 还能进一步引入时序卷积网络(TCN)捕捉长期的舌象演变趋势, 实现疗效与预后的评估, 推动医学与人工智能学科交叉应用的发展。

关键词

舌诊, 深度学习, 图像分析, 卷积神经网络

Research Progress on the Application of Deep Learning Tongue Diagnosis Models in Gastrointestinal Diseases

Yingqing Lin¹, Xirong Liu^{2*}

¹The First Clinical Faculty of Guangxi University of Chinese Medicine, Nanning Guangxi

²Department of Gastrointestinal Endoscopy, The First Affiliated Hospital of Guangxi University of Chinese Medicine, Nanning Guangxi

Received: April 19, 2026; accepted: May 12, 2026; published: May 21, 2026

*通讯作者。

文章引用: 林映清, 刘熙荣. 深度学习舌诊模型在消化道疾病中的应用研究进展[J]. 临床医学进展, 2026, 16(5): 1723-1731. DOI: 10.12677/acm.2026.1651975

Abstract

Tongue diagnosis, as one of the representative methods of inspection in the four diagnostic methods of traditional Chinese medicine, enables doctors to infer internal conditions from external observations. However, traditional tongue diagnosis is often limited by the doctor's experience and subjectivity, thus failing to fully exert its potential. The emergence of various tongue diagnosis models based on deep learning has provided intelligent support for the early screening and syndrome differentiation treatment of upper gastrointestinal diseases. This paper explores the development process of objective tongue diagnosis, retrieves high-quality studies published in the past decade, and conducts a systematic review of the progress in tongue image acquisition and processing, model network architecture, model optimization, and the application research of deep learning tongue diagnosis models in digestive diseases. It is found that existing studies still have shortcomings such as limited sample size and few multimodal models. In the future, we should move towards a multi-center, large-sample, and standardized collection approach. We can combine the tongue diagnosis model with gastroscopy images for comprehensive prediction, thereby improving the prediction accuracy of the intelligent tongue diagnosis model. Additionally, we can further introduce the temporal convolutional network (TCN) to capture the long-term evolution trend of tongue conditions, enabling the assessment of therapeutic effects and prognosis, and promoting the development of interdisciplinary applications between medicine and artificial intelligence.

Keywords

Tongue Diagnosis, Deep Learning, Image Analysis, Convolutional Neural Network

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

消化道癌症日益威胁全球居民身体健康, 2022 年全球胃癌发病率达 4.9%, 食管癌死亡率达 4.6% [1], 2018 年仅中国就有超过 277,000 例食管鳞状细胞癌病例、181,000 例胃癌病例[2], 其中某些消化道病变与癌症密切相关, 例如 Barrett 食管、肠道腺瘤等, 一些消化性溃疡也存在恶变风险。癌前病变是肿瘤在发生、发展、转变过程中的一个重要节点, 早期进行筛查与预防, 对于提高消化道疾病患者生活质量具有重要意义[3], 而这正与中医理论中的未病先防理念不谋而合, 然而传统中医舌诊常受限于医生的经验和主观性而无法充分发挥其作用, 因此, 本文对舌诊客观化发展和深度学习舌诊模型在消化道疾病中的应用进行了探究。

2. 舌诊客观化发展历程

2.1. 舌诊中医理论基础

有诸内, 必形诸外, 祖国传统医学认为舌与脏腑通过经络循行相联系。《黄帝内经》最早描述了不同疾病在舌上不同的反映, 其中《灵枢》有云: “脾足太阴之脉……连舌本、散舌下”, “足厥阴肝经络舌本”, 足见舌象与脏腑联系之紧密。至汉唐时期, 张仲景在《金匱要略》中提到“唇痿舌青”“舌即难言, 口吐涎”等, 将舌诊作为重要的辨病辨证工具; 至元代, 《敖氏伤寒金镜录》的出现奠定了舌诊学的基础, 而后明清时期《温热逢源》中提到“至舌苔之色, 必邪在胃中蒸郁, 其浊气乃上熏而生苔”, 舌

诊体系进一步独立成熟, 表明舌苔乃胃气熏蒸上承于舌面所形成。现代教科书总结前人经验, 将望舌要点概括为以下几点: 望舌神、舌色、舌形、舌态、苔质、苔色等, 舌面的脏腑分候常根据五脏划分, 可分为舌尖、舌中、舌根、舌边四部分, 分别对应心肺、脾胃、肾及下焦、肝胆等。心火亢盛、肺阴不足者, 临床常见舌尖色红、点刺等表现, 脾虚湿困往往表现为舌中淡胖、苔白厚腻等, 舌边瘀斑则通常提示肝郁气滞。由此可见, 舌象的变化可清楚反映脏腑的健康状况。

2.2. 舌诊与现代疾病间的联系

有学者搜索大量文献总结得到诸多与舌苔形成有关的菌群、代谢、蛋白、免疫等生物标志物可以辅助诊病辨证[4], 例如腻苔较薄苔的肠道菌群多样性更高, 证明了舌苔诊病辨证的可靠性。有研究记录经由专家团队辨识的慢性非萎缩性胃炎患者的舌色、舌形等, 与胃镜象(即黏膜粗糙、粘膜白相、粘膜糜烂等表现)联合进行统计分析, 得出结论: 淡红舌和暗红舌在胃粘膜粗糙/非粗糙患者中比例差异有统计学意义, 暗红舌与粘膜粗糙呈正相关性[5]; 但陈中倩却认为: 相较于红或绛舌, 粘膜粗糙与瘀斑瘀点青紫色更具相关性[6]。以上研究均证明了舌象与上消化道疾病的相关性, 两相对比下亦能看出传统中医舌诊中, 以医者肉眼观察舌象特征缺乏标准性, 往往受限于医生的经验和主观性而无法充分发挥其作用。

2.3. 舌象采集与处理标准化

随着线上就诊的普及, 患者自行采集的舌象图片常因曝光、聚焦、雾气等外界因素影响其真实性, 加剧了医生舌诊的困难程度, 因此人们发明出一系列专业仪器用以采集舌象, 舌诊仪的出现使得舌象的采集更加规范, 从 TongueVision 到 TCM-1000 舌诊仪, 再到 Thermal Tongue Diagnosis Device 等, 近年来舌诊仪发展迅速。周明瀚等人将舌诊仪所采集舌象用于研究原发性高血压阴虚阳亢证的舌象特征[7], 江涛使用上海中医药大学团队设计的 TFDA-1 和 TDA-1 舌诊仪采集体检受试者舌象并构建舌诊模型[8], 可见舌诊仪正在被越来越多的人应用以减少深度学习计算量并消除舌体以外区域的干扰。段梦遥等人使用 TFDA-1 将冠心病受试者舌象与健康人舌象相比, 得出冠心病组舌象以舌色偏暗, 苔色偏黄为主的结论[9]; 国外学者将肺癌受试者与良性肺结节患者相比, 发现肺癌组舌色较暗红、苔色则偏黄, 良性肺结节组舌苔厚度大于肺癌组及健康对照组[10]。范宝超等人同样发现肺癌舌苔表现主要为腻苔, 尤其以白腻和黄腻苔为主, 中晚期则多见紫舌、瘀斑和齿痕舌, 胃癌则易出现青紫舌、淡白舌和胖大舌, 舌苔多为白腻苔[11], 以上均证明舌象特征客观化用于预测疾病的可行性。

除了使用舌诊仪进行采集外, 也有学者在舌象处理技术上进行探索, 越来越多学者利用舌象分割技术对舌象图片进行处理, 舌象分割可以去除背景干扰, 保证模型能够准确识别舌象, 传统的舌象分割方法通常需要大量的人力使用 Labelme 标注数据后进行训练, 标注的人工及时间成本较高, 常选用的模型有 Deeplab V3、Deeplab V3+等。此后 Meta 推出了一款视觉分割模型 Segment Anything Model (SAM), 该模型在图像识别任务中表现更佳, 该模型通过预训练的深度学习网络来进行图像分割, 通过自监督学习训练, 不依赖于大量人工标注的数据集, 使得模型的泛化能力更强, 能够适应不同种类的图像分割任务[12]。且 SAM 采用了基于 Transformer 的结构, 通过自注意力机制使分割的准确性和灵活性增强以识别更复杂的图像特征, 其交互式分割的特点使得由于分割导致部分舌体缺失的概率大大下降[13], 研究者仅需通过点击或框选来辅助模型进行分割, 此后仅需筛选出小部分未达标舌象, 再次进行人工标注后分割, 大大减少了人力及时间成本, DeeplabV3 [14]在这方面的灵活性则弱于 SAM。

然而, 仅对舌象的采集与处理进行规范化远远不够, 不同医师对舌象的判断仍存在一定的差异性, 因此在舌象采集与处理技术逐渐规范的同时, 人们将舌诊与深度学习融合, 通过现代数据处理技术来量化舌象, 使舌诊一致性进一步提升。深度学习舌诊较传统舌诊在诊断一致性、诊断效率、适用范围、动

态记录能力方面均具有较大优势, 因此深度学习模型逐渐成为了大规模舌象特征提取的首选工具, 具体如表 1 所示。多名学者利用深度学习模型识别齿痕舌[15]、裂纹舌[16] [17]等均取得良好结果。

Table 1. Differences between traditional tongue diagnosis and deep learning-based tongue diagnosis
表 1. 传统舌诊与深度学习舌诊的区别

对比维度	传统舌诊	深度学习舌诊
一致性	不同医师、同一医师在不同阶段的诊断一致性均可能有较大差别, 受医师临床经验、当下看诊状态以及周围环境影响较大, 主观性强。	无论是不同医师、不同时间或是不同地点, 诊断一致性较高, 成熟模型可稳定输出, 不受外界干扰, 客观性良好。
诊断效率	青年医师需经过多年临床学习才能够积攒舌诊心得, 平均单个病例诊断需数分钟完成, 应对大规模筛查较为艰难。	经专人训练成功后, 仅需简单培训即可熟练使用, 降低技术门槛, 单个病例诊断耗时数秒至 1 分钟不等, 可大批量分析。
适用范围	较为局限, 仅适用于个体化辨证论治, 受地区资源分布限制较大。	可应用于大规模筛查、远程医疗等。
动态记录能力	通过病历书写记录及医生记忆回顾不同阶段舌象变化, 难以精准对比, 即使使用手机、相机等拍照设备留存照片, 储存及后续搜索较为麻烦。	可储存历次标准化舌象图片, 并输出特征量化数据, 可直观对比舌象变化趋势。

3. 深度学习舌诊模型

3.1. 网络架构

深度学习网络有卷积神经网络(CNN)、循环神经网络(RNN)、生成对抗网络(GAN)和 Transformer 等, 其中最适合处理图像数据的当属 CNN [18], CNN [19]是一类广泛应用于图像处理与模式识别领域的深度学习模型, 主要由卷积层、池化层及全连接层构成, 经预处理的数据集通过卷积层, 卷积层使用不同尺寸的卷积核(如 3×3 、 7×7)对舌象进行滑动卷积提取图像中的局部特征(如轮廓、舌苔纹理、舌色分布等)后通过池化层, 池化层一般使用最大池化或平均池化, 可以在降低特征空间维度的同时保留关键特征信息, 减少计算量并防止过拟合, 最后将高层特征图转变为一维向量, 通过全连接层即可完成图像的最终分类或回归任务。它能够从自动学习输入的图像数据中的高级特征, 浅层网络通常仅能够提取边缘、纹理和颜色等较为表浅的信息, 深层网络则进一步学习目标的局部结构、整体轮廓及语义特征, 因此在图像分类、目标检测和医学图像分析等任务中表现出较强的特征表达能力, 常见的经典 CNN 有 AlexNet、VGGNet、GoogLeNet、ResNet 等。

AlexNet 由 Alex Krizhevsky 等人在 2012 年提出的一种 CNN 网络架构[20], 它不仅采用了 ReLU 激活函数替代传统函数, 加快了训练速度, 还使用最大池化保留显著特征, 避免模糊化效果。此外, 该网络首次使用多 GPU 进行计算, 使网络能向更深层次发展, 同时还引入了 Dropout 机制用以大幅减少过拟合现象, 提高模型泛化能力。但因其相对于后续出现的网络来说网络层数仍相对较浅, 参数量主要集中在全连接层, 模型较为庞大, 因此特征提取能力和参数利用效率相对较差。

VGGNet [21]由牛津大学视觉几何组于 2014 年提出的经典深度卷积神经网络之一, 它通过反复堆叠小卷积核代替大卷积核, 结合 ReLU 函数和最大池化层, 结构清晰地同时又使感受野扩大。该网络虽证明了使用较小的卷积核增加网络深度可以提高模型效果, 但参数量仍较大, 仍占用较多存储和计算资源, 残差网络与其相比更为高效深入。

GoogLeNet 是 2014 年由 Google 提出的一种经典卷积神经网络[22], 是深度学习图像分类发展过程中的一个重要代表模型。它由若干卷积层、池化层和多个 Inception 模块串联组成, 核心特点就是引入了 Inception 模块, 传统卷积神经网络在同一层通常只使用一种卷积核大小, 而 GoogLeNet 的 Inception 模块

则并行使用 1×1 、 3×3 、 5×5 卷积核, 再将这些不同分支的输出在通道上拼接起来, 即可同时提取不同尺度的特征。其中, 较小的卷积核适合捕捉局部细节, 较大的卷积核适合提取大范围的语义信息, 二者协同增强了模型的多尺度特征表达能力。此外, 该网络中间位置设有辅助分类器, 它能够在训练阶段使中间层更好地学习到有效特征, 减少梯度传播困难, 起到一定的正则化作用。但该模型也有一定的局限性, 由于并行分支较多较为复杂, 不如残差网络统一, 因此在迁移学习方面通常不如后续更新的网络灵活。

ResNet 是 2015 年由微软研究院的何凯明[23]提出的里程碑式的一种网络架构, 其核心是引入了残差连接, 有效减少了深层网络梯度消失和网络退化问题, 使 CNN 能够向更深层次迈进。其中经典代表模型之一是 ResNet50, 其网络层数顾名思义共有 50 层, 由多个残差块堆叠而成, 不仅能够提取低级纹理且能读取高级语义, 该网络不仅结构清晰、训练稳定, 而且迁移能力较强。但该模型由于采用单路径的残差结构, 在特征表达多样性方面相对于后续的引入多分支设计的 ResNeXt-50 ($32 \times 4d$)来说就较为有限。

ResNeXt-50 ($32 \times 4d$) [24]在保留 ResNet 框架的基础上引入了分组卷积, 该网络将每个残差块内的卷积划分为 32 个执行相同变换的分支, 每个分支内有 4 个卷积核, 然后将这些分支的输出进行聚合, 在不显著增加参数数量的前提下提升特征表达, 因此该网络的特征提取能力较以上网络更强。

以上各类模型均具有各自的优缺点, 具体对比如表 2 所示。

Table 2. Comparison of five convolutional neural networks

表 2. 五种卷积神经网络对比

	AlexNet	VGGNet	GoogleNet	ResNet50	ResNeXt-50 ($32 \times 4d$)
出现时间	2012 年	2014 年	2014 年	2015 年	2017 年
结构特点	首次使用 ReLU 函数; 引入了 Dropout; 双 GPU 训练; 全部采用最大池化	使用小卷积核, 增加层级堆叠	含 Inception 模块 (多尺度卷积), 在同一层内多分支同时提取特征	深层网络结构, 通过残差连接解决深层网络的梯度消失和退化问题	分组卷积 + 多分支结构, 在 ResNet 基础上引入“基数”概念, 通过多个并行分支增强特征表达能力
优势	加快了训练速度, 最大池化保留显著特征, 双 GPU 使网络能向更深层次发展	扩大感受野, 证明使小卷积核增加网络深度可提高模型效果, 有较强的泛化能力	参数量较少, 计算效率较高, 适合较轻量的模型	能够训练更深的网络, 特征表达能力较强, 结构稳定	在参数量可控的情况下多分支计算, 提升特征表达能力, 性能通常优于 ResNet
局限性	网络层数仍相对较浅, 模型较为庞大, 特征提取能力和参数利用效率较为有限	参数量仍较大, 仍占用较多存储和计算资源。	结构较复杂, 模块设计较为固定, 对深层特征表达能力有限	网络深度增加后计算量较大, 特征提取能力仍受单路径卷积限制	结构相对复杂, 计算资源需求较高

3.2. 模型优化

在 CNN 的研究与应用中, 为提升模型的预测性能与泛化能力, 研究者通常从数据处理、类别不平衡处理、正则化策略、调整参数等多个方面进行改进, 以此来提高模型在复杂任务中的稳定性与准确性。

为了提升模型的性能, 在数据层面, 研究者们常对数据进行数据增强处理。常见的数据增强方式有几何变换, 水平翻转[20]、裁剪、缩放、遮挡[25]、擦除[26]等都属于几何变换, 能够增强模型对图像空间变换的适应能力。除此之外, 还有颜色空间增强, 即通过调整图像的亮度、对比度、饱和度等提高模型在不同光照条件下的鲁棒性。数据增强通过对原始训练样本进行一系列变换, 使数据集中生成新的训练

样本而增加数据多样性, 进而使模型的泛化能力有所提升。

在深度学习任务中我们常常会遇到类别不平衡问题, 类别不平衡指模型在训练过程中容易偏向样本数量较多的类别, 而对少数类别的识别能力较差的情况, 诸多学者们针对该问题探索出了一些数据平衡方法对模型进行调整。数据平衡可以通过增加少数类别的样本量或减少多数类别样本的影响来令模型, 主要包括过采样、欠采样及类权重调整等。过采样方法通过生成或复制少数类样本以扩大其规模, 如 SMOTE [27]、ADASYN 等, 欠采样方法则通过减少多数类样本数量来实现类别平衡, 主要包括随机欠采样和基于聚类的欠采样方法, 聚类欠采样如 K-means 聚类、C 均值聚类等。其中, SMOTE 通过在少数类样本及其近邻样本之间进行线性插值生成新的合成样本, 从而扩展少数类特征空间; ADASYN 方法则在 SMOTE 基础上根据样本的分布自适应生成样本, 对少数类分布区域生成更多合成样本。随机欠采样通过随机删除部分多数类样本降低其比例, 而基于聚类的欠采样方法则通过对多数类样本进行聚类并选取代表性样本, 从而在减少样本数量的同时尽可能保留原始数据分布特征, 提高模型训练的稳定性及泛化能力。类权重调整[28]则是通过在损失函数中引入类别权重, 令模型更关注少数类别, 但如果权重选择不当, 可能会使模型对少数类别过于敏感, 忽视多数类别的关键特征, 反而降低了模型的泛化能力。从以上论述我们了解到各种数据平衡处理方法均有各自的优缺点, 于是有些学者提出了将上述各类平衡方法相互结合的想法, 例如 Qiao Xu 等人在构建深度学习模型中创新性地提出了 BSFCM 混合采样算法(即将 borderline-SMOTE 算法和模糊 C 均值聚类下采样算法结合) [29], 有效减少了数据类之间及数据类之内的不平衡。

此外, 为了减少损失函数的值, 诸多研究者不断优化算法用以调整模型参数, 例如随机梯度下降法 (SGD)、动量法、Adam、RMSprop 等。随机梯度下降法在一次次迭代中计算梯度, 并沿着下降梯度的反方向更新参数, 该优化算法较为基础, 但收敛速度较慢易受局部震荡影响, 往往需要多次调整。动量法则增加了动量项, 在过去梯度的加权平均, 能够加速收敛并减少震荡。Adam 优化器结合了动量法和自适应学习率的优点, 可以通过自动调整学习率来提高模型训练的稳定性及收敛效率。RMSprop 则通过对梯度平方的指数加权平均来动态调整学习率, 能够适应不同参数的更新需求。

为了防止模型在训练过程中出现过拟合, 人们引入了多种正则化方法, 常见的正则化方法例如 Dropout [30], 该方法利用在网络训练过程中随机丢弃部分神经元来减少模型对特定特征的依赖; 又如 Batch Normalization, 其通过归一化使网络中各层特征的分布稳定, 从而加快训练收敛、提高训练稳定性, 具有一定的正则化效果[31]; 除此之外还有早停策略[32], 当模型在验证集上的性能不再提升时早停策略能够令模型及时停止训练, 这样做不仅能够减少验证误差, 还能节省计算资源。

综上所述, 在卷积神经网络的研究中, 模型性能的提升通常依赖于多方面技术的综合应用。从数据增强与数据质量控制, 到类别不平衡处理, 再到模型参数优化、正则化策略等方面的改进, 都大大提升了深度学习模型性能。在具体应用中, 根据任务特点和数据特征选择合适的优化策略, 并对不同方法进行合理组合, 往往能够显著提升模型的预测能力与泛化性能。

4. 研究应用成果

Ma C 等人采用 YoLoV5 模型对舌象进行分割后, 用舌象及胃癌风险指标(年龄、性别、Hp 感染情况等)进行预测, 根据病理诊断将受试者进行二分类, 分为癌前病变(PLGC)组和非癌前病变(非 PLGC)组, 将预测结果和根据病理诊断分类结果进行相关性分析, 腻苔、裂纹舌、暗舌、黄苔、厚苔这 5 个舌象标签具有统计学意义[33], 这在一定程度上说明了通过舌象发现风险特征的医学意义, 对癌前病变有一定的评估作用。此外, 该团队还发现幽门螺杆菌感染与 PLGC 和非 PLGC 呈弱相关, 而现有共识认为 Hp 感染是胃癌最突出的危险因素[34] [35], 幽门螺旋杆菌与癌前病变及胃癌的关系值得进一步研究。

Li Yuan 等人进行了一项多中心、观察性的研究, 招募了 10 个中心的胃癌与非胃癌参与者, 分别训练了 DeepLabV3+、APINet、transferg 三种模型, 并另外招募 7 个中心的胃癌(GC)和非胃癌(NGC)参与者作为验证集进行外部验证。三个模型均表现出对舌象图像的判别有效性, GC 患者的舌象表现为舌苔厚度明显增加、水润度明显下降, 这表明舌象可以作为 GC 诊断的稳定工具[36], 美中不足的是该研究未将糜烂、溃疡、肠上皮化生以及高级别上皮内瘤变等胃相关疾病纳入研究设计中。

Xiaohe Sun [37]等人用收集处理了 1389 张结直肠癌患者的舌象图片和 1543 张非结直肠癌受试者的舌象图片, 使用 Grounding DINO 结合 SAM 来实现舌体图像识别分割, 并基于 Swin-Transformer 构建了舌诊模型, 结果在内部验证中准确率 87.93%, F1 分数高达 0.9072。此外, 他们还进行了独立外部验证, 结果得出准确率: 85.18%, 召回率: 85%, F1 分数: 0.8507, 各性能评价指标结果十分可观, 证明了该舌诊模型用于结直肠癌辅助筛查具有巨大潜力。

张景慧等人[38]收集了健康人群、Hp 感染、反流性食管炎、胆汁反流、胃和十二指肠糜烂等受试者的舌象图片共 3140 张, 构建舌诊模型, 得到总体 AUC: 0.808, 准确率: 0.753, 灵敏度: 0.850, 特异度: 0.683, 该模型纳入疾病种类较为丰富, 训练时间成本低, 且再次证明了舌象与胃肠道疾病之间有紧密联系, 通过舌诊结合深度学习算法预测上消化道疾病十分具有探索意义。

5. 结语

综上所述, 舌诊作为中医诊断的重要组成部分, 在深度学习算法的辅助下能够有效地从舌象中提取特征并与疾病信息关联, 尤其是可以实现对癌前病变的评估, 在癌症的筛查与预防中展现出巨大潜力。未来的研究应尽量扩大样本量, 聚焦于提高舌诊模型的准确性和泛化能力, 同时应尽可能与其他医学检查结合, 例如黄丽等人做了一项利用胃镜图像构建良恶性胃溃疡辅助诊断系统模型的回顾性研究[39], 徐伟超等人亦构建了胃镜图像自动识别模型, 经训练后系统准确性可达 98% [40], 美中不足的是该研究为单中心回顾性分析, 未来可尝试进一步扩大样本量, 针对罕见病或特定证型样本不足的问题, 可引入生成对抗网络(GAN)生成高质量合成舌象数据, 在扩充数据集的同时提升模型对少见特征的识别能力, 最终将舌诊模型与胃镜图像综合进行预测, 形成综合诊断体系。此外, 在此类模型成熟的基础上可进一步引入时序卷积网络(TCN)捕捉长期的舌象演变趋势, 实现疗效与预后的评估。

参考文献

- [1] Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., et al. (2024) Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, **74**, 229-263. <https://doi.org/10.3322/caac.21834>
- [2] Arnold, M., Ferlay, J., van Berge Henegouwen, M.I. and Soerjomataram, I. (2020) Global Burden of Oesophageal and Gastric Cancer by Histology and Subsite in 2018. *Gut*, **69**, 1564-1571. <https://doi.org/10.1136/gutjnl-2020-321600>
- [3] 张绍丽, 曹毛毛, 杨帆, 等. 上消化道癌前病变患者健康相关生活质量评价及影响因素研究[J]. *中国肿瘤*, 2024, 33(9): 747-755.
- [4] 杨闪闪, 娄彦妮, 贾立群. 舌苔形成机制的研究进展[J]. *中华中医药杂志*, 2022, 37(10): 5857-5860.
- [5] 李萌, 于靖文, 丁媛, 等. 慢性萎缩性胃炎患者舌象、脉象与胃镜象相关性分析[J]. *辽宁中医杂志*, 2024, 51(5): 10-14.
- [6] 陈中倩. 慢性萎缩性胃炎唇象、舌象、胃镜象辨证分布规律研究[D]: [硕士学位论文]. 济南: 山东中医药大学, 2017.
- [7] 周明瀚, 刘旺华, 李花, 等. 原发性高血压阴虚阳亢证舌象客观化研究[J]. *中华中医药杂志*, 2022, 37(6): 3401-3404.
- [8] Jiang, T., Lu, Z., Hu, X., Zeng, L., Ma, X., Huang, J., et al. (2022) Deep Learning Multi-Label Tongue Image Analysis and Its Application in a Population Undergoing Routine Medical Checkup. *Evidence-Based Complementary and Alternative Medicine*, **2022**, Article ID: 3384209. <https://doi.org/10.1155/2022/3384209>

- [9] 段梦遥, 王楚皓, 谈宇权, 等. 315 例冠心病患者舌象特征客观化研究[J]. 中医杂志, 2024, 65(9): 921-927.
- [10] Shi, Y., Guo, D., Chun, Y., Liu, J., Liu, L., Tu, L., *et al.* (2023) A Lung Cancer Risk Warning Model Based on Tongue Images. *Frontiers in Physiology*, **14**, Article 1154294. <https://doi.org/10.3389/fphys.2023.1154294>
- [11] 范宝超, 黄旭晖, 谭为. 肿瘤舌象信息研究进展[J]. 世界科学技术-中医药现代化, 2020, 22(5): 1614-1618.
- [12] Dosovitskiy, A., Beyer, L., Kolesnikov, A., *et al.* (2021) An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale. arXiv: 2010.11929.
- [13] Mazurowski, M.A., Dong, H., Gu, H., Yang, J., Konz, N. and Zhang, Y. (2023) Segment Anything Model for Medical Image Analysis: An Experimental Study. *Medical Image Analysis*, **89**, Article ID: 102918. <https://doi.org/10.1016/j.media.2023.102918>
- [14] Chen, L.C., Papandreou, G., Schroff, F., *et al.* (2017) Rethinking Atrous Convolution for Semantic Image Segmentation. arXiv: 1706.05587.
- [15] Tang, W., Gao, Y., Liu, L., Xia, T., He, L., Zhang, S., *et al.* (2020) An Automatic Recognition of Tooth- Marked Tongue Based on Tongue Region Detection and Tongue Landmark Detection via Deep Learning. *IEEE Access*, **8**, 153470-153478. <https://doi.org/10.1109/access.2020.3017725>
- [16] Yan, J., Cai, J., Xu, Z., Guo, R., Zhou, W., Yan, H., *et al.* (2023) Tongue Crack Recognition Using Segmentation Based Deep Learning. *Scientific Reports*, **13**, Article No. 511. <https://doi.org/10.1038/s41598-022-27210-x>
- [17] Hu, J., Yan, Z. and Jiang, J. (2022) Classification of Fissured Tongue Images Using Deep Neural Networks. *Technology and Health Care*, **30**, 271-283. <https://doi.org/10.3233/thc-228026>
- [18] 卢宏涛, 张秦川. 深度卷积神经网络在计算机视觉中的应用研究综述[J]. 数据采集与处理, 2016, 31(1): 1-17.
- [19] Lecun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998) Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, **86**, 2278-2324. <https://doi.org/10.1109/5.726791>
- [20] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2017) ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, **60**, 84-90. <https://doi.org/10.1145/3065386>
- [21] Simonyan, K. and Zisserman, A. (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv: 1409.1556.
- [22] Szegedy, C., Liu, W., Jia, Y., *et al.* (2014) Going Deeper with Convolutions. arXiv: 1409.4842.
- [23] He, K., Zhang, X., Ren, S., *et al.* (2015) Deep Residual Learning for Image Recognition. arXiv: 1512.03385.
- [24] Xie, S., Girshick, R., Dollár, P., *et al.* (2017) Aggregated Residual Transformations for Deep Neural Networks. arXiv: 1611.05431.
- [25] DeVries, T. and Taylor, G.W. (2017) Improved Regularization of Convolutional Neural Networks with Cutout. arXiv: 1708.04552.
- [26] Zhong, Z., Zheng, L., Kang, G., *et al.* (2017) Random Erasing Data Augmentation. arXiv: 1708.04896.
- [27] Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P. (2002) SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research*, **16**, 321-357. <https://doi.org/10.1613/jair.953>
- [28] Ling, C. and Sheng, V.S. (2008) Cost-Sensitive Learning and the Class Imbalance Problem. In: Sammut, C. and Webb, G.I., Eds., *Encyclopedia of Machine Learning*. <https://link.springer.com/referencework/10.1007/978-0-387-30164-8>
- [29] Qiao, X., Lu, C., Duan, M., Liu, Z., Liu, Y., Chen, W., *et al.* (2024) Intelligent Tongue Diagnosis Model for Gastrointestinal Diseases Based on Tongue Images. *Biomedical Signal Processing and Control*, **96**, Article ID: 106643. <https://doi.org/10.1016/j.bspc.2024.106643>
- [30] Srivastava, N., Hinton, G., Krizhevsky, A., *et al.* (2014) Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, **15**, 1929-1958.
- [31] Ioffe, S. and Szegedy, C. (2015) Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. arXiv: 1502.03167.
- [32] Prechelt, L. (2012) Early Stopping—But When? In: Montavon, G., Orr, G.B. and Müller, K.R., Eds., *Neural Networks: Tricks of the Trade*, Springer, 53-67. https://doi.org/10.1007/978-3-642-35289-8_5
- [33] Ma, C., Zhang, P., Du, S., Li, Y. and Li, S. (2023) Construction of Tongue Image-Based Machine Learning Model for Screening Patients with Gastric Precancerous Lesions. *Journal of Personalized Medicine*, **13**, Article 271. <https://doi.org/10.3390/jpm13020271>
- [34] 顾晓, 邢莹莹. 幽门螺杆菌 CagA 诱导胃部炎症癌转化的研究进展[J]. 中国药科大学学报, 2025, 56(1): 132-138.
- [35] 赫捷, 陈万青, 李兆申, 等. 中国胃癌筛查与早诊早治指南(2022, 北京) [J]. 中华肿瘤杂志, 2022, 44(7): 634-666.
- [36] Yuan, L., Yang, L., Zhang, S., Xu, Z., Qin, J., Shi, Y., *et al.* (2023) Development of a Tongue Image-Based Machine

-
- Learning Tool for the Diagnosis of Gastric Cancer: A Prospective Multicentre Clinical Cohort Study. *eClinicalMedicine*, **57**, Article ID: 101834. <https://doi.org/10.1016/j.eclinm.2023.101834>
- [37] Sun, X., Huang, L., Qu, L., Chen, C., Zeng, X., Zhou, Z., *et al.* (2025) Development of a Tongue Image-Based Machine Learning Tool for the Diagnosis of Colorectal Cancer: A Prospective Multicentre Clinical Cohort Study. *IEEE Journal of Biomedical and Health Informatics*. <https://doi.org/10.1109/jbhi.2025.3585552>
- [38] 张景慧, 王娟, 赵玉洁. 基于机器学习的胃肠道疾病舌诊模型构建[J]. 山东大学学报(医学版), 2024, 62(1): 38-47, 70.
- [39] 黄丽, 李艳霞, 吴练练, 等. 基于深度学习的良恶性胃溃疡人工智能辅助诊断系统研究[J]. 中华消化内镜杂志, 2020, 37(7): 476-480.
- [40] 徐伟超, 李博林, 许亚培, 等. 基于快速区域卷积神经网络萎缩性胃炎-胃癌胃镜图像自动识别模型的建立及临床测试[J]. 世界科学技术-中医药现代化, 2021, 23(9): 3274-3280.