

深度学习在舌象分类中的研究综述

吴 欣¹,徐 红^{1,2},林卓胜¹,李胜可¹,刘慧琳³,冯 跃¹⁺

1. 五邑大学 智能制造学部,广东 江门 529020

2. 维多利亚大学,澳大利亚 墨尔本 8001

3. 上海中医药大学 教学实验中心,上海 201203

+通信作者 E-mail: j002443@wyu.edu.cn

摘要:随着技术的快速发展和计算能力的提高,深度学习在舌象分类领域将得到广泛的应用。舌诊图像的舌象分类是中医舌诊客观化的重要组成部分。传统的舌诊是在基础理论的指导下,借助个人经验所做出的理解和判断,因而会具有一定的差异性和模糊性,影响诊断的可重复性。为了减少主观判断的误差,许多研究人员致力于通过深度学习实现中医舌诊的客观化、定量化和自动化。主要对基于深度学习的舌象分类方法研究现状进行分析梳理和归纳总结。在舌象分类研究中,以各类深度学习方法作为研究对象,将其划分为基于早期神经网络、卷积神经网络、区域卷积神经网络、迁移学习以及其他方法进行总结分析;对舌诊中的中医证候和疾病以及体质分类进行了讨论;用Kaggle上的公开舌诊数据集进行5折交叉验证实验,数据集为小样本齿痕舌,评估了基于深度学习和迁移学习分类方法;对舌诊图像质量、构建数据集方式、特征提取、单标签和多标签分类的研究发展进行了探讨和展望。

关键词:中医舌诊;深度学习;舌象分析;舌象分类

文献标志码:A **中图分类号:**TP18; TP391.41

Review of Deep Learning in Classification of Tongue Image

WU Xin¹, XU Hong^{1,2}, LIN Zhuosheng¹, LI Shengke¹, LIU Huilin³, FENG Yue¹⁺

1. Faculty of Intelligent Manufacturing, Wuyi University, Jiangmen, Guangdong 529020, China

2. Victoria University, Melbourne 8001, Australia

3. Experiment Center of Teaching & Learning, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China

Abstract: With the rapid development of technology and the improvement of computing power, deep learning will be widely used in the field of tongue classification. The classification of tongue image is an important part of tongue diagnosis in traditional Chinese medicine (TCM). Traditional tongue diagnosis is dependent on understanding and judgment skills gained from personal experience under the guidance of basic theory, which leads to ambiguity and variability, affecting diagnostic reproducibility. In order to reduce the error of subjective judgment, many researchers have devoted themselves to realizing the objectification, quantification and automation of tongue diagnosis in TCM

基金项目:广东省普通高校重点领域专项(2021ZDZX1032);广东省国际及港澳台高端人才交流专项(2020A1313030021);五邑大学科研项目(2018TP023,2018GR003)。

This work was supported by the Basic Research Key Project in General Colleges and Universities of Guangdong Province (2021ZDZX1032); the International and Hong Kong-Macao-Taiwan High Talent Exchange Special Project of Guangdong Province (2020A1313030021); the Scientific Research Project of Wuyi University (2018TP023, 2018GR003).

收稿日期:2022-08-15 **修回日期:**2022-10-24

through deep learning. This paper analyzes and summarizes the research status of tongue image classification methods based on deep learning. In the study of tongue image classification, various deep learning methods are used as the research objects. The research objects are divided into categories based on early neural networks, convolutional neural networks, regional convolutional neural networks, transfer learning and other methods for summary analysis. TCM syndromes/diseases in tongue diagnosis and classification of physical constitution are discussed. A 5-fold cross-validation experiment is conducted with the public tongue diagnosis dataset on Kaggle. The dataset is a small sample of the dentate tongue, and the classification methods based on deep learning and transfer learning are evaluated. The research development of single-label and multi-label classification is discussed and prospected.

Key words: traditional Chinese medicine tongue diagnosis; deep learning; tongue image analysis; tongue image classification

中医历经几千年的发展,已形成了一个集诊断、治疗、预防于一体的医疗体系。作为中医诊断方法之一的舌诊的理论基础在于舌的异常变化,如纹理、颜色和形状的变化,可揭示人体对应器官的疾病,诊断易行,无疼无损^[1-3]。中医临床研究表明,在疾病的发展过程中,舌的变化是迅速而显著的,然而传统的舌诊有赖于医者的主观分析,有不可避免的局限性^[1],诊断结果可能是模糊和不确定的,阻碍了其发展应用。因此,有必要对舌诊进行客观化定量检测,并探讨舌象与疾病的量化关系。

如今,计算机科学和技术的发展使得基于图像处理和模式识别的自动舌诊系统设计成为可能^[4]。机器学习算法已被应用于舌象的颜色、纹理和形状研究,包括支持向量机(support vector machines, SVM)^[5]、随机森林(random forest, RF)^[6]、K-均值聚类(K-means clustering algorithm)^[7]、主成分分析法(principal component analysis, PCA)^[8]、梯度提升决策树(gradient boosting decision tree, GBDT)^[9]、自适应提升(adaptive boosting, AdaBoost)^[10]和朴素贝叶斯(naive Bayes model)^[14]等。然而,被采集者的舌象特征复杂且相互关联,如舌色与苔色间的相关性较为明显,要达到预期的分类精度具有一定的难度。

传统的图像分类算法通常在简单的图像分类任务上效果较好,但在较为复杂的任务上效果差强人意^[11-13]。随着人工智能的快速发展,深度学习技术也越来越成熟,智能方法应用于中医研究^[14-17]使该领域发生了范式转变。刘飞等人^[18]基于深度学习的方法将预处理的图像特征映射作为网络的输入,完成了对糖尿病患者和正常对照组的分类识别。在临床应用上,不少研究者开展舌诊以及相关领域客观化和定量化的探索,如Sun等人^[19]基于7层卷积神经网络

(convolutional neural networks, CNN)提出了一种无需边界框准确分类齿痕舌并定位舌诊图像中重要区域的方法,平均分类准确率达到78.6%。

分析统计目前已有的舌诊研究文献发现(见图1^[20]),文献发表数量整体呈波浪式上升趋势,但增加速度缓慢,研究主题主要包括舌诊客观化研究、中医舌诊、中医舌象、中医诊断、图像分割、图像处理、特征提取和冠心病等内容^[20]。研究方法主要是通过传统的机器学习和深度学习对舌诊图像进行定量分析。传统方法适合解决小样本场景中过度拟合的问题。基于深度学习在图像分类方面取得较好的成果,因此一些深度学习的方法被广泛应用于舌诊客观化研究。深度学习与传统方法相比具有以下优点:(1)在大规模数据和识别精度上,深度学习模型具有更好的能力;(2)可以最大程度地进行信息交换。

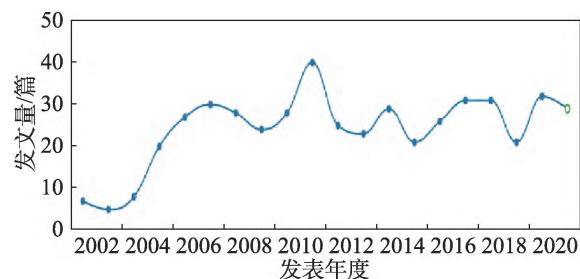


图1 舌诊客观化研究发表文献的变化趋势

Fig.1 Growth of objective study papers on tongue diagnosis

目前,舌诊自动化主要的处理流程分析(见图2^[21])如下:首先采集舌诊图像,进行颜色校正或数据增强(平移、旋转和镜像等)预处理;然后基于深度学习进行舌体分割和舌象分类,舌象分类包括提取舌象的颜色、纹理和形状等特征与舌象分析;接着根据中医舌诊知识库的相关知识实现中医证候的推断;最后给出中医诊断建议。

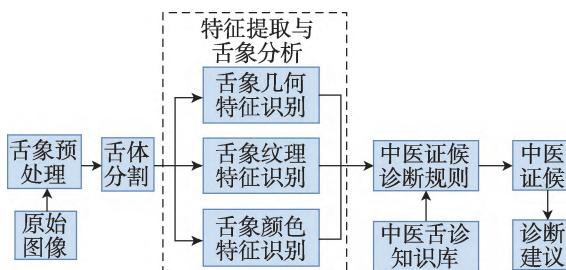


图2 舌诊自动化流程

Fig.2 Tongue diagnosis automated process

1 基于深度学习的舌象分类方法

1.1 早期神经网络

人工神经网络由大量相互连接的神经元组成,它近似模拟了神经信号的处理。Rosenblatt等人提出了多层感知器模型(multilayer perceptron, MLP),首次将神经网络的理论付诸实践^[22-24]。Obafemi-Ajayi等人^[13]设计了自动识别和舌形分类的方法,同时用SVM和MLP来训练舌形分类,并提出了基于对称轴检测的新方法来校正舌头的方向。在自建的包含303张舌象的数据集上的实验结果表明,MLP比SVM识别准确率高约3%。

1998年,LeCun等人^[25]提出基于梯度学习的CNN模型LeNet-5,网络由卷积区和全连接(fully connected, FC)区两部分组成。邢甜甜^[26]对舌色、舌苔色、舌苔薄厚、舌苔腐腻和舌苔润燥5个类别分别建立了CNN模型,参照LeNet-5网络进行调整。同时,建立了多任务CNN模型,将颜色和纹理分类任务分别对应一个网络模型。与单任务分类相比,在节约时间、减少计算量的同时不牺牲分类性能。

1.2 卷积神经网络

LeNet-5网络逐渐被其他机器学习方法所超越^[27-29]。2012年,Krizhevsky等人^[30]构建了AlexNet网络。Zeiler和Fergus^[31]构建ZFNet网络,使用可视化技术揭示神经网络各层的作用并优化网络,调整后的网络性能超越了原始网络。随后,原理、算法和模型经历了井喷式发展,相关的研究工作如汗牛充栋。

Xue等人^[32]用裂纹舌和非裂纹舌数据对AlexNet网络进行训练,提取裂纹区域的高层次特征。Huo等人^[33]在AlexNet网络基础上,通过批量归一化(batch normalization, BN)^[34]对网络进行优化,结果表明预处理方法提高了舌形分类的精度,减少了训练时间,证明了对不同舌形识别的有效性。肖庆新等人^[35]对AlexNet结构进行修改,将卷积核大小为 11×11 或 7×7 改为 5×5 或 3×3 ,并且减少了全连接层的输出

数量,卷积操作后增加了池化操作,减少了参数数目和特征图尺寸,搭建了舌苔颜色分类模型。

针对2012年的ImageNet数据集,深度学习框架Caffe定义的结构与AlexNet大致相同,被称为CaffeNet^[36]。Hou等人^[37]针对舌色分类问题,在CaffeNet的基础上对网络结构进行修改,修改方案与文献[33]相同。因为舌象数量有限,在学习过程中,舌象的某些特征会逐渐减弱甚至消失,所以减少FC层的神经元数量,结果比传统方法更实用、准确。

2014年,Simonyan等人^[38]提出了VGG模型,运用更小的卷积核,并且加深网络,一定程度减少了参数数量,达到更好的效果。AlexNet、VGG等模型都是通过增大网络的深度来获得更好的训练效果,但层数的增加会带来很多副作用,比如过拟合、梯度消失/爆炸等。2015年,Szegedy等人^[39]结合文献[40-41]提出GoogLeNet网络,Inception模块的提出可以从另一种角度优化训练效果,在相同计算量下提取到更多的特征,更高效地利用计算资源。

为了解舌苔的性质,Fu等人^[42]采用Otsu阈值分割^[43]舌体,并将图片色彩转化到HSV^[44]颜色空间,由于大部分舌苔位于舌的中央或根部,且在水平方向上近似对称,采用了水平翻转的数据增强方式,最后用GoogLeNet网络进行了四分类和三分类。CNN具有局部性和平移不变性等优点。但如果需要对图像中的对象或错误进行定位、标定、代码读取或后期处理,则传统算法更具有优势。该方法结合了基础图像处理和深度学习的特点,使得模型性能有所提升。

CNN在图像分类任务中表现良好。然而,随着CNN深度的增加,训练更加困难,训练误差更高。随着网络深度的增加,准确率达到饱和后迅速下降。2015年,He等人^[45]提出了ResNet解决这种退化问题。为减少中医师之间的诊断差异,Li等人^[46]提出一种基于CNN的舌特征分类框架(见图3),使用一种改进的人脸标志点检测方法和U-Net^[47]完成舌分割任务,最后以ResNet-34作为骨干网络,从舌象中提取特征并分类。邵尤伟^[48]先将舌质与舌苔分离,然后将分离后的图像作为输入,使用ResNet-50分别对舌质和舌苔分类,并参考Zhang教授的方法^[49]预测病症和诊断结果。Chang等人^[50]针对裂纹舌建立ResNet-50和Grad-CAM^[51]可视化裂隙区域,该模型既可检测舌裂纹,又可定位裂纹区域。当提取固定位置的手工特征时,舌苔区域的大小或位置变化可能会导致模型性能不一致。为了解决这一问题,Tang等人^[52]利用ResNet对舌苔特征进行提取和采用多实例支持向量机

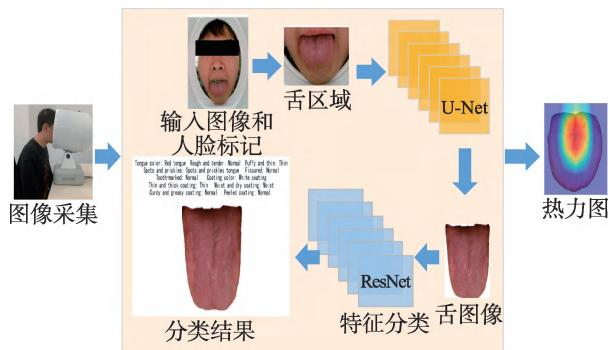


图3 舌象分类整体框架

Fig.3 Overview of tongue classification framework

(multiple-instance support vector machines, MI-SVM)^[53]进行分类。

孙萌等人^[54]将 ResNet 网络与 Inception 模块^[39]采用 NIN (network in network)^[41]方式结合,设计了 Triplet Loss 的度量分类方法(见图 4),最大化非同类样本的特征距离,同时缩小类间样本特征的间距,增加了网络的深度与宽度,同时能更好地提取特征,减少参数量。

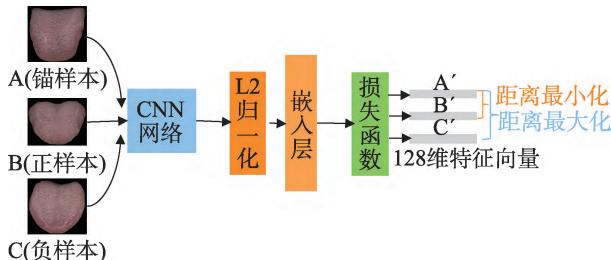


图4 算法整体框架

Fig.4 Overall framework of algorithm

残差连接使CNN更深、更强、更高效。DenseNet^[55]进一步扩展网络连接确保各层间的信息流最大化。陈慧贞^[56]以 AlexNet 网络结构为基础,引入 DenseNet 中的密集连接方式,提出一种舌苔分类模型 TonNet。又因为舌诊需舌苔、舌质等多方位综合诊断,研究人员以 TonNet 为基础上设计了包含舌苔和舌质两个输出的舌象辨识模型 Multi-TonNet。

目前舌象分类中注意力机制使用最多的是 SENet (squeeze-and-excitation network)^[57]。牛富泉^[16]对包含人脸的舌图像进行分割,得到的舌区域图像采用带有 SE block 的 ResNet-101 网络分类,提高了裂纹舌的识别准确率。为提高智能舌诊的诊断效果和增加舌诊的可解释性,骆明楠^[58]提出基于 Competitive-SE 通道注意力机制的 CNN 和随机区域池化的训练方式,提升了表征能力,突出了某些重要特征,并根据舌象信息提供诊断思路(见图 5),从而促进舌诊的可解释性

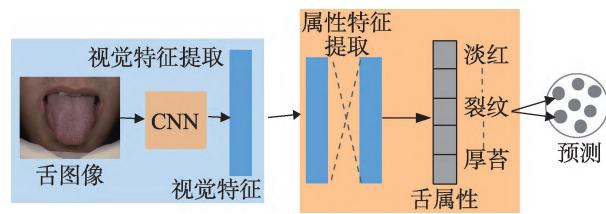


图5 基于舌属性学习的舌诊预测

Fig.5 Prediction of tongue diagnostic based on tongue attribute learning

和智能化。

1.3 基于区域卷积网络的模型

自 2014 年 R-CNN (regions with CNN features)^[59]问世以来,目标检测领域涌现大量改进的模型^[60],例如 Fast R-CNN^[61]、Faster R-CNN^[62]。齿痕区域仅出现在舌边缘区域。现有的方法^[63-64]大多是利用凹区域信息对齿痕舌进行分类,当齿痕区域不凹时,分类结果往往不一致。为了解决这些问题,Li 等人^[65]提出三阶段的方法:第一阶段,用 R-CNN 选择搜索生成舌象可疑区域;第二阶段,将可疑区域输入 CNN 提取每个区域的特征向量;第三阶段,CNN 的输出被视为一个二维矩阵,并使用多实例分类器进行最终决策。Gholami 等人^[66]利用 R-CNN 网络将舌体区域与其他区域分离,经过必要的预处理后,将图像提供给 DenseNet 网络分类。

杜春慧^[67]针对舌图像的裂纹、颜色和齿痕做了定量研究,并搭建了中医舌诊推荐系统,一定程度上为医生和患者提供了便利。研究者采用 Faster R-CNN 网络识别裂纹,通过 ZFNet、VGG-16 和 ResNet-101 提取裂纹特征。同时采用模糊 C 均值聚类分离舌质和舌苔,通过 KNN、SVM、CNN 和 Bayes 识别舌色。Chen 等人^[68]提出基于 ResNet-50 模型和注意力机制改进的 Faster R-CNN 方法。首先,考虑到随着网络深度的增加,缺失的信息会增加,提出在区域候选网络 (region proposal network, RPN) 中使用 ResNet-50 模型,以保留更多的原始信息。其次,在网络中引入空间注意力和通道注意力机制,强调局部病变信息,抑制背景舌象信息。结果与原始的 Faster R-CNN 相比,准确率显著提高。

Mask R-CNN^[69]在 Faster R-CNN 基础上增加分支目标掩码预测网络,并行当前的边界框识别,能同时进行目标检测和分割,且易于扩展到其他任务上。Li 等人^[70]将数据分为数据集 1 和数据集 2,并利用 Mask R-CNN 模型对数据集 2 的舌诊图像进行检测,去除原始图像的背景。然后在数据集 1 和数据集 2 上采用

叠加技术进行模型融合,以最大限度地发挥临床数据的优势,获得更好的分类效果。Peng 等人^[71]设计了舌象点刺识别与提取模型,包括多尺度特征图生成、候选区选择和目标区域识别。利用深度卷积网络构造多尺度特征图,然后利用目标候选框生成算法和选择策略选择高抽象的特征图,最后利用分类网络对目标区域分类。在非标准光照条件下,实验使用手机拍摄各种舌象,以增加模型的鲁棒性。

根据舌象纹理的特点,刘佳丽等人^[72]设计基于YOLOv3 目标检测改进的 Double-D 优化架构。首先在特征提取架构 DarkNet-53 网络中进行细粒度特征融合,然后利用 K-means 维度聚类和多尺度识别定位,得到训练权重后,对舌裂纹进行预测。Yen 等人^[73]使用 YOLOv4^[74]对舌所属的类别进行分类并获取坐标信息,可用于识别其他舌象。针对齿痕舌识别效率低和准确率不高的问题,颜建军等人^[75]提出基于 YOLOv5 的齿痕检测模型,利用分割算法对舌体进行分割,采用 YOLOv5 检测齿痕区域,提取相关齿痕舌特征向量,用 RF 分类。

Weng 等人^[76]提出了基于 DarkNet-53 的弱监督 YOLO(weakly supervised YOLO,WSYOLO)方法(见图 6),利用完全边界框注释和粗图像注释的舌图像来训练齿痕和裂纹检测模型。该模型基于 YOLO 扩展而来,并增加了几个分类分支,用于识别齿痕舌和裂纹舌。分类分支从不同层次提取特征,检测分支用于定位齿痕和裂纹的位置。

1.4 迁移学习

中医舌诊需要观察舌体的颜色、纹理和形状,以获得必要的信息。由于医学图像样本少,采用迁移学习方法缓解了医学图像样本不足的问题。目前基于迁移学习的舌象分类方法大多是利用大型 ImageNet 数据集或其他医学数据集对网络预训练,然后迁移到目标任务对网络微调。最后使用少量医学数据重新训练全连接层进行分类,进而提高网络的性能。医学图像的主流迁移框架如图 7 所示,其中微调的模型、次数、顺序以及预训练数据集的选用都会影响最终迁移的效果^[77]。

针对开放环境中拍摄引起的舌体位置不统一问题,李渊彤等^[21]用预训练的 VGG-16 模型初始化,用空间变换层纠正图像目标的空间位置,再用多个卷积层和池化层的组合获取多个任务之间共享的语义特征图,最后由一个卷积块和多个全连接层组合用于共同的语义特征提取,针对各自任务相关的语义特征向量进行分类预测。

Chen 等人^[78]提出了相似稀疏域自适应方法对舌诊进行建模(见图 8),其中 A 表示 ResNet-34 网络,B 表示 Faster R-CNN。在 ResNet-34 预训练的基础上构建 RPN 的训练,只需增加少量计算成本。通过对输入的舌图像进行分类,将具有代表性的病变和其他异常标记出来作为最终多分类的辅助依据,有效地模拟中医师诊断。

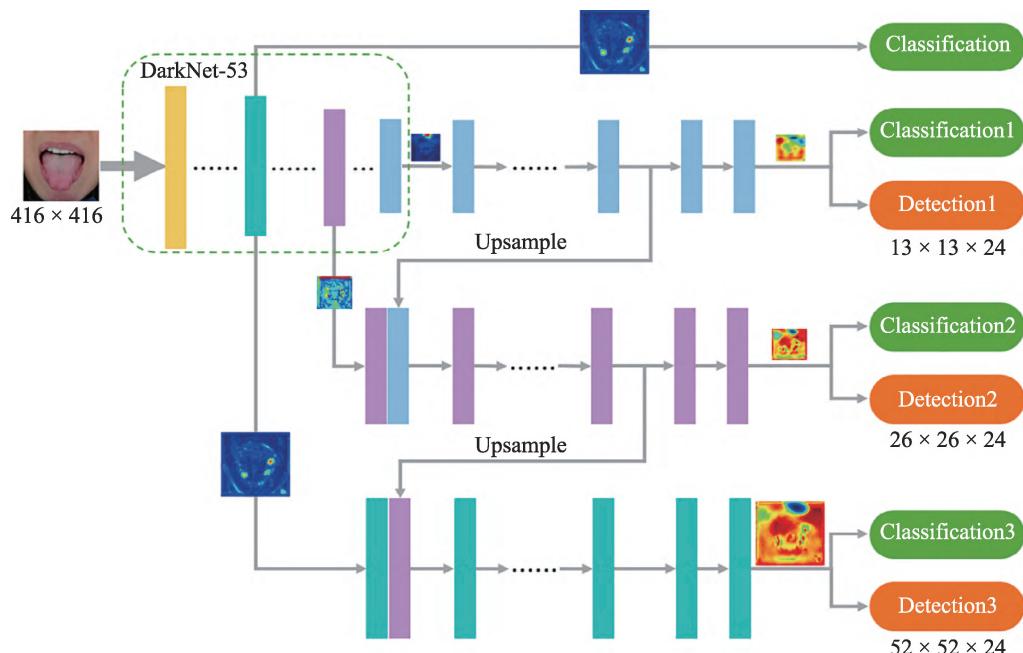


图 6 WSYOLO 的网络架构

Fig.6 Framework of WSYOLO

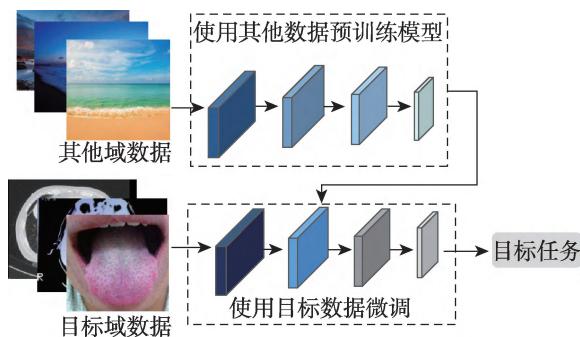


图7 医学图像的主流迁移框架

Fig.7 Mainstream migration framework in medical image field

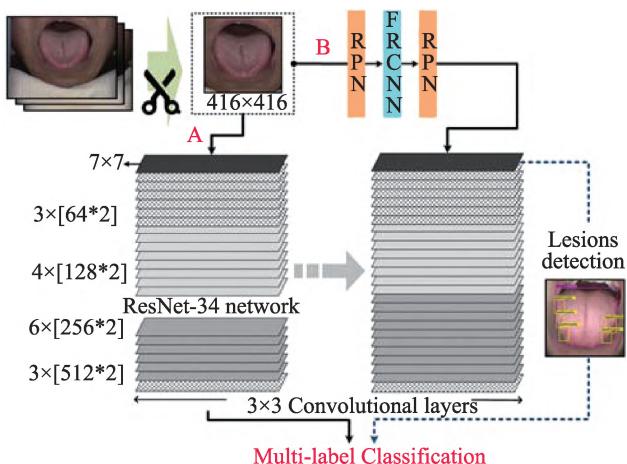


图8 中医舌诊模型架构

Fig.8 Architecture of TCM tongue diagnosis model

刘婧玮^[79]基于 ResNet-34 网络和 VGG-16 网络对齿痕和腻苔分类,通过迁移学习和数据增强优化防止过拟合,但因为样本不平衡且腻苔诊断受医生主观影响较大,腻苔识别在验证集上的准确率不高。Wang 等人^[80]用预训练权重初始化网络并微调,使用 ResNet-34 网络结构提取特征并分类。该模型可推广到其他设备捕获的图像,具有较强的泛化能力,可从信息学的角度提供客观、方便的计算机辅助舌诊方法,用于跟踪疾病进展和评估药理作用。

邱童^[81]用 ResNet-50 预训练并微调,采用 Faster R-CNN 框架对舌诊图像进行预处理,使用多特征融合的方法对高低语义信息特征融合,对是否是舌象进行分类。为了适应各种摄影环境下的舌诊图像,构建中药处方,Hu 等人^[82]用 ResNet-50 和 VGG 模型预训练,并设计了用于处方构建的神经网络框架。此外,研究人员还提出辅助治疗主题丢失机制,以模拟中医师治疗,并减少输出标签对结果多样性的干扰。实验使用真实的舌诊图像和相应的处方,结果可

以生成与真实样本接近的处方,验证了方法的可行性。同时,从更多的物理信息出发,为自动构建中药处方提供参考。Song 等人^[83]通过预训练好的 ResNet 和 Inception-v3 网络提取舌体特征,用全局平均池化和全连接层重写原始网络的输出层,得到分类结果。

Sadasivan 等人^[84]采用经典的 AlexNet、VGG、GoogLeNet 和 ResNet 模型进行迁移学习,并采用两种方法进行对比:第一种方法,在大规模的 ImageNet 数据集上,使用微调的预训练模型,最佳准确率为 98.61%;第二种方法,使用未微调的预训练模型,准确率为 96.94%。尽管与第二种方法相比,第一种方法的训练时间较长,但研究结果强调迁移学习在识别任务中的优越性,并且 ResNet 模型效果都优于其他模型。

杨晶东等人^[85]利用迁移学习获取公共特征,再使用全连接神经网络对舌象特征进行训练分类,架构采用“Inception-v3 和两层全连接层”和“Inception-v3 和三层全连接层”。为解决舌象分类易受到舌部其他信息干扰的问题,翟鹏博等人^[86]提出融合注意力机制的多阶段舌象分类算法,并设计了舌象采集系统采集舌诊图片。首先通过融合不同的感受野特征,获得舌部区域,然后借助注意力机制排除舌部其他的信息干扰,提取更为精准的信息进行分类。因为有的被采集者齿痕不明显且与周围舌体较为接近,所以齿痕舌比其他类别的分类准确率低。

齿痕舌的识别具有一定的挑战。现有的方法大多只关注齿痕舌的分类,没有涉及齿痕的确切位置和数量,对后续诊断缺乏指导性作用。Kong 等人^[87]试图解决这些问题,提出了一种基于 Mask scoring R-CNN 框架和在 COCO 数据集上训练 ResNet-101 网络的迁移学习方法,可以提取、可视化齿痕并识别齿痕的数量。该方法可以稳定地检测和分割齿痕,为齿痕舌的严重程度分析提供依据。刘梦等人^[88]以带预训练权重的 VGG-16 和 ZFNet 为特征提取器的 Faster R-CNN 模型对齿痕舌和裂纹舌进行局部特征提取。

由于舌体分割的复杂性和细粒度特性,舌诊图像分类具有一定的挑战,而两项任务是高度相关的,与多任务联合学习(multi-task joint learning, MTL)的思想相同。Xu 等人^[89]提出了用于舌诊图像分割和分类的 MTL 方法。其中,网络在 ImageNet 数据集上进行预训练,以 ResNet-50 为网络的骨干,并且考虑到舌苔分类的类间方差小、类内方差大的特点,同时将 U-Net 和判别滤波器学习融合到 MTL 中,使这两项任务能达到预期的效果。

舌象标签之间有较强的相关性,需综合考虑舌象多标签。汤一平等^[90]提出了MTL模型。首先,在ImageNet数据集上参数初始化,在共享网络层对所有标签进行联合学习,自动挖掘和利用标签间的相关性;然后,在不同子网络层分别完成特定类别的学习任务,从而消除多标签分类中的歧义性;最后,训练多个Softmax分类器以实现对所有标签的并行预测。该方法以端到端的方式同时提取舌象多个特征并直接进行分类识别。

1.5 其他方法

舌象标签之间有较强的相关性,需综合考虑舌象多标签。王爱民等人^[91]提出了基于学习矢量量化(learning vector quantization, LVQ)神经网络的舌色、苔色自动分类方法,分类器采用“ 2σ ”准则进行样本筛选,并采用Fisher^[92]作为色度空间(RGB, HIS, Ohta, YUV)选择的依据,有效地提高了图像的识别率。然而,一个图像块可能有多种类别的彩色,但在标注时仅将其定为一种类别,容易造成量化误差。

文献[93]在图像模糊相关求和过程中提出了加权因子的算法,然而这是以人工方式完成的。针对这个问题,Qin等人^[94]提出模糊小脑模型关节控制器(fuzzy cerebellar model articulation controller, FCMAC)取代加权模糊关联算法,以克服参数调整过程中的不确定性和不适应性。首先,根据专家的判断选择图像作为输入,训练FCMAC神经网络进行模糊适应,并作为检验对象。其次,提高期望输出值以增加主要特征颜色向量的比例,同时降低次要特征向量的比例,以提高对检索焦点的关注度。人眼对颜色具有相对恒常性,而机器不敏感颜色的变化,通过加入噪声,如阴影或日光图,降低表示阴影和日光图的颜色向量的比例,以便检索对噪声不太敏感。

深度学习网络通常使用大量的舌诊图像进行训练。然而,某些类别的舌图像很少,这使得深度学习无法使用。此外,收集大量的舌诊图像成本太高且难以应用。Qiu^[95]采用元学习方法,通过几个样本提高识别的准确性,提出了改进的原型网络(prototypical network),并根据舌诊图像的特点设计了基于领域知识(domain knowledge)和混合损失函数的特征嵌入算法。该方法比改进前的原型网络、ResNet和AlexNet的精度都有提升,具有良好的泛化能力。

鉴于多模型决策可以克服单模型判别的不足,获得更高的分类精度。Xiao等人^[96]提出了多模型决策的中医舌苔纹理分析方法。首先,对舌苔纹理进行预分类。然后,通过比较Inception-v3、ResNet-50和MobileNet等不同深度网络模型的分类精度,进行

多模型决策,进一步优化舌苔纹理分析结果。

通用CNN对区域舌象细节特征的获取能力较弱,同时忽略卷积通道之间的关系,导致模型有较高冗余。为增强CNN效果,Hu等人^[97]提出了随机区域合并方法获取详细的区域特征。同时,还提出了内成像通道关系建模方法,并与空间注意力机制相结合,对所有通道上的多区域关系建模。新的全通道区域注意力网络对局部细节特征建模,提高了建模效率。该模型为舌象细节特征的建模提供了范例,对其他辅助诊断应用具有指导意义。关于如何动态地调整任务的交互时间,Xu等人^[98]提出了任务耦合弹性学习框架建立任务相关性,用特征图像对疾病位置和疾病性质分类,用动态序列模块建立序列关系并实现多阶段训练和损失正则化。

采集到的原始数据通常是有噪声的高维数据,Dai等人^[99]提出了概念对齐深度自动编码器的方法分析舌图像。深层自动编码器分解为两个主要部分:编码器逐层执行对输入图像的降维,解码器重建原始图像。Meng等人^[100]提出了特征提取框架,以提取无偏差的特征并减少舌诊断的工作量。为解决冗余问题,引入高分散性和局部响应正常化操作。针对中医舌色分类中存在噪声的问题,卓力等人^[101]提出有噪声标注下的舌色分类方法(见图9):首先,用二阶段的数据清洗;然后,用基于通道注意力的轻量卷积神经网络增强特征的表达能力;最后,用知识蒸馏策略提高分类性能,教师网络用ResNet-50和SENet网络,学生网络用轻量CNN网络。此外,Rajakumaran等人^[102]提出了深度可分离卷积模型用于舌色特征提取与分析,在预处理阶段包括数据增强和双边滤波去除噪声。利用舌图像数据集对该模型的性能进行评估,结果表明具有良好的分类性能。

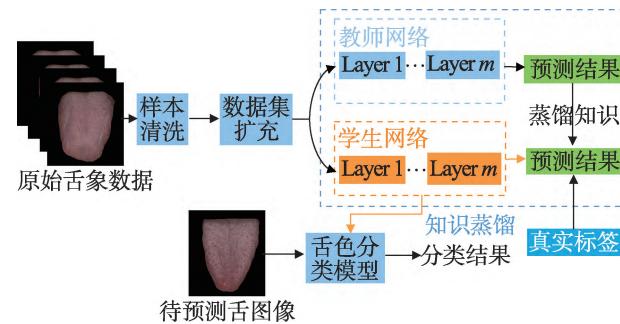


图9 有噪声标注的中医舌色分类方法整体框架

Fig.9 Overall framework of Chinese medicine tongue color classification method with noisy labeling

早期神经网络结构简单且深度较浅,适用于单

一旦无复杂背景的舌体图像;卷积神经网络增加了网络深度或宽度等,提升了特征提取能力,更关注舌体图像局部区域,适用于开放和封闭环境下的大部分舌象;区域卷积神经网络可以实现疾病位置、齿痕和裂纹定位,适用于背景复杂且明显的舌体图像;迁

移学习能够缓解样本不足的问题,降低计算资源损耗,缩短训练时间,适用于小样本舌象分类;其他方法是针对某一特定问题提出的解决方案。

综上所述,深度学习算法在中医学中的应用发展迅速,表1和表2总结了深度学习的舌象分类方法。

表1 基于深度学习的舌象分类方法比较

Table 1 Comparison of deep learning based methods for tongue classification

模型	文献	主要网络结构	优点	缺点	整体归纳
早期神经网络	[13]	MLP	采用对称轴检测校正舌方向	忽略细节信息	网络结构简单,适用于单
	[26]	LeNet	建立多任务和单任务的CNN模型	网络结构没有大的变化	一旦无复杂背景的舌象
	[32][33][37]	AlexNet	关注的是舌局部区域,加入BN优化	需要选择合适的数据集 获取参数	
	[43]	GoogLeNet	使用HSV模型	与二分类器相比,多分类器在某些类别上的准确性略有下降	
	[46]	ResNet-34	对舌体图像的特征向量提取和分类,而不是对整个舌区提取和分类	分割效果影响分类结果	
卷积神经网络	[48]	ResNet-50	舌质与舌苔分离/可视化裂隙区域	未考虑细节信息	
	[54]	Triplet Loss度量分类	最大化非同类样本的特征距离同时缩小类间样本特征的间距	增加网络的复杂性	网络训练时间长,参数多,表征能力强,适用于开放和封闭环境下的舌象
	[56]	DenseNet	增强特征传播,加强特征复用	增加训练参数和网络的复杂性	
	[16]	SE Block ResNet-101	增强有用的信息,抑制无用的信息	未考虑舌象特征复杂且相互的关联性	
	[58]	Competitive-SE	提取更多的细节特征信息,提升诊断效果	人为的标注可能导致属性标注不够全面或不够准确	
	[65][66]	R-CNN	生成可疑区域并提取每个区域的特征向量	候选区域高度重叠,冗余计算太多	
	[67][68]	Faster R-CNN	保留更多的原始舌体信息	时间消耗大	
	[69][70]	Mask R-CNN	易于扩展到其他任务	结构无变化	
区域卷积神经网络	[72]	YOLOv3	考虑了舌象细微复杂的特点	输入图像大小需要结合具体数据分析	可实现疾病位置、齿痕和裂纹的定位,适用于背景复杂且明显的舌体图像
	[73]	YOLOv4	实现了速度和精度的平衡	仅判断是否是舌象	
	[75]	YOLOv5	灵活性与速度较好	舌边分割效果直接影响检测齿痕区域的准确度	
	[76]	WSYOL0	从不同层次提取特征,定位了齿痕和裂纹的位置	未考虑细节信息	
	[91]	LVQ	采用“ 2σ ”准则进行样本筛选	标注时仅将它定为一种类别,易造成量化误差	
	[94]	FCMAC	克服了参数调整过程中的不确定性和不适应性	需要加入噪声,增加模型的鲁棒性	
	[95]	Prototypical Network	适用于小样本数据集	需要选择合适的数据集 获取参数	
其他方法	[96]	MobileNet	先预分类再多模型决策分析	未考虑舌象特征复杂且相互的关联性	适用于解决特定问题的舌象分类
	[97]	内成像通道关系建模	提出内成像通道注意力机制、动态区域共享机制	增加网络的复杂性	
	[98]	任务耦合弹性学习	考虑了任务相关性	需要大量数据	
	[99]	自动编码器	概念对齐深度	信息易丢失	
	[101]	知识蒸馏	增强特征的表达能力	网络结构没有变化	
	[102]	深度可分离卷积模型	采用数据增强和双边滤波的去除噪声	对特定的数据有效	

表2 基于迁移学习的舌象分类方法比较

Table 2 Comparison of transfer learning based methods for tongue classification

文献	源数据	目标数据	网络模型	实验结果/%
[21]	ImageNet	528张临床舌象数据	VGG-16	几何特征平均准确率:82.58; 宏查准率:88.87;宏查全率:84.34 舌象纹理平均准确率:82.07; 宏查准率:87.53;宏查全率:84.23
[78]	地面裂纹	1 500张舌病变裂纹	类似稀疏域适应、ResNet-34、Faster R-CNN	类别平均达到80以上,最高可达91.60
[79]	ImageNet	1 760张临床舌象数据	ResNet-34、VGG-16	齿痕识别准确率:90.50 灵敏度:87.25;特异性:1.28
[80]	ImageNet	1 548张临床舌象数据	ResNet-34	准确率:90.50±1.60 灵敏度:87.20±3.29 特异性:93.00±1.28
[81]	ImageNet	网上收集的185张舌图片	ResNet-50、Faster R-CNN	准确率:86.49;灵敏度:86.36;特异性:90.48
[82]	ImageNet	958张临床舌象数据及其相应的处方	ResNet-50和VGG	处方构建的召回相似度:37.17
[83]	ImageNet	2 245张临床舌象数据	ResNet、Inception-v3	齿痕识别准确率:96.88 裂纹识别准确率:94.00 舌苔厚度识别准确率:96.88
[84]	ImageNet	900张临床裂纹舌数据	AlexNet、VGG、GoogLeNet和ResNet	准确率 AlexNet:98.33;VGG-16:96.11 VGG-19:96.66;GoogLeNet:97.50 ResNet:98.61
[85]	ImageNet	网上收集的180张舌图片	Inception-v3	准确率 Inception_v3+2CNN:90.30 Inception_v3+3CNN:93.98
[87]	COCO	1 500张临床舌象数据	ResNet-101	精度:99.00 召回率:91.40 F1-score:95.00
[88]	模型参数迁移	600张临床舌象数据	Faster R-CNN	宏观准确率:91.00;宏观精确率:97.70;宏观召回率:86.00
[89]	ImageNet	1 858张临床舌象数据	ResNet-50	F1-score:93.02
[90]	ImageNet	1 222张临床舌象数据	MTL	(基于图像)精确度:98.94;召回率:99.07;F1-score:98.92; 杰卡德相似度:98.29 (基于类别)精确度:98.33;召回率:98.45;F1-score:98.37; 杰卡德相似度:96.88

2 常用数据集和评价指标

2.1 常用数据集

对于中医舌象分类来说,由于中医临床数据的特殊性和隐私保护问题,可研究的数据有限,目前的公开数据集较少。公开数据集 Tooth-Marked-Tongue 为 HANHUI 在 Kaggle 网站上发布的尺寸不一的有齿痕和无齿痕数据,共 1 250 例^[103],如图 10 所示,其舌象的标签是由多名中医师标注。公开数据集 BioHit 图像数据集^[104]包含 300 例舌图像,图像尺寸为 567 × 768 像素,如图 11 所示。目前大多数研究者自建数据集,从互联网收集,与各个学术研究机构和医院合作,并由专业且经验丰富的中医专家标注,如图 12 所示。

由于医学图像样本数量不足,在图像分类之前,需对样本进行数据增强,以保证不同特征的舌象数

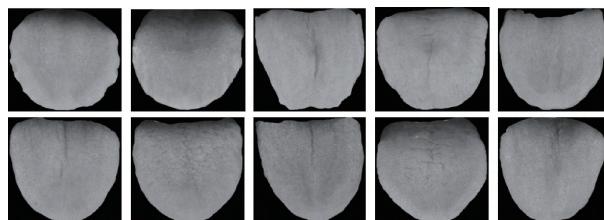


图10 Tooth-Marked-Tongue 数据集部分舌象图像

Fig.10 Some samples of Tooth-Marked-Tongue dataset

据具有相同规模的训练集和验证集。采用的类别平衡策略为保持最大的样本数目不变,再根据少数类与最大类的数量间的差异,对少数类进行样本类别扩充。常用的数据增强方式有平移、旋转、水平翻转、加高斯噪声等不同变换方法的组合来对图像进行增强,也有利用生成模型生成新的样本或新的特征。



图 11 BioHit 数据集部分舌象图像

Fig.11 Some samples of BioHit tongue dataset



图 12 自建数据集部分舌象图像

Fig.12 Some samples of self-built tongue dataset

SMOTE (synthetic minority oversampling technique)^[105] 是一种综合采样人工合成数据的算法, 用于解决数据类别不平衡问题, 以过采样少数类和下采样多数类结合的方式来合成数据。Qi 等人^[6]利用 SMOTE 算法解决舌色学习样本数的不平衡性来提高分类精度。

2.2 评价指标

无论是二分类问题, 还是多分类问题, 最常用的是性能度量指标是错误率和准确度。然而, 只有错误率和准确度不能满足所有任务的要求, 往往还有查准率(precision, P)与查全率(recall, R)。

对于二分类问题, 可依据真实类别与预测类别划分为真正例(true positive, TP)、假正例(false positive, FP)、真反例(true negative, TN)、假反例(false negative, FN), 如表 3 所示。

表 3 混淆矩阵

Table 3 Confusion matrix

真实类别	预测类别	
	正例	反例
正例	TP (真正例)	FN (假反例)
反例	FP (假正例)	TN (真反例)

准确度(accuracy, Acc)是指分类正确的样本数除以样本总数, 由式(1)计算:

$$Acc = \frac{TP + TN}{TP + TN + FN + FP} \quad (1)$$

查准率 P 和查全率 R 分别如式(2)和式(3)所示:

$$P = \frac{TP}{TP + FP} \quad (2)$$

$$R = \frac{TP}{TP + FN} \quad (3)$$

一般来说, 查全率 R 低时, 查全率 P 往往高, 反之亦然。

F1 度量是基于查准率与查全率的调和平均, 由式(4)计算:

$$F1 = \frac{2 \times P \times R}{P + R} \quad (4)$$

特异性(specificity, Spc)是指负样本被预测为负样本的比例, 计算方式如式(5):

$$Spc = \frac{TN}{FP + TN} \quad (5)$$

虚警率(false alarm rate, FAR)是指负样本中被分为正样本的比例, 其值越小说明模型的性能越好, 计算公式如式(6):

$$FAR = \frac{FP}{FP + TN} \quad (6)$$

舌象分类大多都为单标签分类, 一般只是针对一个类别, 如颜色、纹理或者形状, 并没有考虑到多标签分类。而病人的情况是复杂多样的, 故诊断任务也有图像多分类任务。舌诊预测任务是数据不平衡下的多标签分类任务, 因此评价指标不能采用单标签图像分类的标准, 要综合考察查准率和查全率。

AP 值为根据精准率和召回率绘制而成的 PR 曲线下的面积, 能够很好地在医疗任务上对模型进行评判。而 mAP 值是多类别的 AP 值, 是衡量模型在多个类别上的预测性能情况, 计算公式如式(7)所示。

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (7)$$

一种比较直接的方法是在各混淆矩阵上分别计算查准率和查全率, 然后计算平均值, 即可得到宏查准率(macro-P)、宏查全率(macro-R)和对应的宏平均(macro-F1), 计算公式如式(8)~(10)所示:

$$macro - P = \frac{1}{n} \sum_{i=1}^n P_i \quad (8)$$

$$macro - R = \frac{1}{n} \sum_{i=1}^n R_i \quad (9)$$

$$macro - F1 = \frac{2 \times macro - P \times macro - R}{macro - P + macro - R} \quad (10)$$

也可以把各个混淆矩阵对应的元素进行平均, 再基于平均值计算微查准率(micro-P)、微查全率(micro-R)和微平均(micro-F1), 计算公式如式(11)~(13)所示。

$$micro - P = \frac{\sum_{i=1}^N TP_i}{\sum_{i=1}^N TP_i + \sum_{i=1}^N FP_i} \quad (11)$$

$$micro - R = \frac{\sum_{i=1}^N TP_i}{\sum_{i=1}^N TP_i + \sum_{i=1}^N FN_i} \quad (12)$$

$$micro - F1 = \frac{2 \times micro - P \times micro - R}{micro - P + micro - R} \quad (13)$$

3 舌诊图像的应用

近年来,图像分类、舌象提取、舌体或舌苔分割等重要的人工智能技术被广泛应用于舌诊,以提高诊断的鲁棒性和准确性。在中医临床实践中,中医师通过观察舌色、形状、舌苔、唾液量等特征,推断出被采集者的主要疾病。不同综合症或疾病的舌象被电子化后,得到标有不同综合症或疾病的舌象数据集,应用人工智能方法构建预测人体不同综合症或疾病状态的舌象,为临床综合症和疾病的诊断及临床用药提供了依据^[106],舌诊图像与中医证候和疾病的映射关系如图13所示。刘国正^[107]将舌象与中医证候联系起来,采用ResNet-18模型进行了4组分类(正常-证候、正常-热证、正常-寒证和热证-寒证),并扩展到3分类(正常-热证-寒证)。

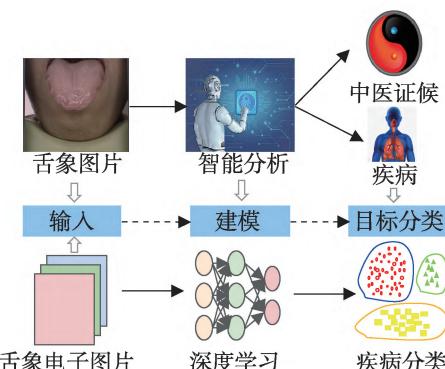


图13 图像与中医证候和疾病的映射方案

Fig.13 Mapping scheme of images to TCM symptoms and diseases

人工智能是通过将表示学习与复杂的推理相结合的系统实现的^[108]。Gui等人^[109]利用标签分布特征进行舌色的多标签学习,通过颜色标签来描述舌象。这种方法可以看作是不同医生对同一样本的诊断结果的总结,符合中医诊断的实际情况。人工智能辅助舌诊可以将2型糖尿病、早期乳腺癌、冠心病等患者与健康人区分开来^[4,79,110-112]。

研究表明糖尿病患者的舌苔常呈黄色且苔厚^[113]。Zhang等人^[114]使用舌色、纹理和几何特征检测糖尿病和非增殖性糖尿病视网膜病变。Li等人^[115]从舌诊图像中提取颜色特征、纹理特征和舌苔,采用最新的深度学习技术ViT(vision transformer)^[116]进行舌象分类,通过Grad-CAM方法提供定位诊断信息,并验证聚类结果的可靠性。ViT的训练速度远优于卷积神经网络和循环神经网络^[117-118]。此方法为中医药预防和治疗糖尿病提供诊断依据,开展个体化治疗;促进

中医诊断的规范化发展;保证中医诊断的准确性和一致性。

随着对冠心病预测系统的最新研究,分类变得非常重要^[119],深度学习提高了冠心病预测系统的准确性。叶桦等人^[120]建立基于反向传播神经网络的糖尿病合并冠心病舌象和脉象的证型预测模型。

2019年冠状病毒病(corona virus disease 2019, COVID-19)引起了全世界的关注。Liang等人^[121]首次将舌诊应用于COVID-19疾病进展的时间序列分析。通过定期分析获得各种舌诊图像,然后根据舌头的特征来调整中药配方。研究人员发现舌色、舌苔厚度和舌苔颜色与COVID-19的病变密切相关。Wang等人^[122]在ImageNet数据集上用ResNet构建了腻舌苔识别网络(GreasyCoatNet)来完成COVID-19疾病舌诊图像分类任务。研究者们用普通摄像机拍摄了50名处于恢复期的COVID-19患者和43名无COVID-19病史的对照组舌象,用GreasyCoatNet34模型量化与COVID-19疾病相关的舌苔。此研究有助于治疗选择和预测,可以为区分舌象、诊断中医证候和跟踪疾病进展(如COVID-19)提供重要的研究范式,在临床应用中显示出独特的潜力。

由于综合症的诊断受限于临床医生的经验,Hsu等^[113]提出了综合诊断系统框架,已实现定量和客观诊断,同时辅助年轻医生准确诊断。研究人员使用多箱检测器(multi-box detector)检测舌体图像的裂缝区域,同时用自己开发的G-TongueNet和L-TongueNet从整个舌体图像和裂缝区域提取全局和局部特征,再使用GBDT对发热证、血虚和脾虚进行分类,确定最终的综合症。

中医体质(traditional Chinese medicine constitution, TCMC)作为中医理论的重要组成部分之一,引起了越来越多的研究人员的关注。国医大师王琦院士团队提出的九种体质学说^[122-123]是通过中医的视角来区分人体生理特征个体差异的概念。

Zhou等人^[124]采用AlexNet网络和灰度共生矩阵,以及最小封闭矩形和边缘曲线来提取舌图像的特征。然后使用不同的分类器对九种体质中的气虚质、痰湿质和湿热质分类,神经网络的分类精度比传统机器学习算法的准确率提高了9%。Zhou等人^[125]继续在AlexNet网络的基础上,搭建了特征融合的深度网络,通过多个特征层的融合,实现对舌诊图像的三种体质分类,得到较好的分类结果,准确率达到了73%,比未使用融合的方法^[124]提高了10%。谢佳澄^[126]

对舌象进行TCMC分类,通过机器学习和深度学习判别一张图片中是否存在舌,用户通过手机上传舌象图片便可进行TCMC分类的判别,并根据体质结果提供饮食、生活方式等相关信息的推荐。Li等人^[127]提出在开放环境下基于混合深度学习的识别方法,采用GoogLeNet网络识别体质。现有的体质识别算法对舌诊图像只给出一个体质结果。然而,在实际的体质认定过程中,个人往往会有两个或两个以上的体质。Li等人^[128]采用多标签分类方法来识别舌诊图像的复合构成,并比较了四种不同网络(Inception-v3、ResNet-50、VGG-16和DenseNet-121)的特征提取性能。

舌诊图像的质量是构建舌诊领域标准数据集的基础。Jiang等人^[129]首先对舌诊图像质量进行评估(image quality assessment, IQA),然后进行颜色校正和舌体分割,对舌色、舌质和舌形提取特征并分析,综合舌诊图像特征得到九种TCMC之一,最后根据具体的体质制定中医药治疗方案。IQA的质量筛选模型,用于识别质量好和质量差的舌诊图像,通过准确度、精密度、召回率和F1-score等指标进行比较。

此外,尽管深度学习方法已经应用于医学领域^[130-131],大多数研究者还没有使用深度学习方法自动从大量训练样本中学习特征。Ma等人^[132]提出了自动体质识别的系统框架,该框架由舌诊图像采集、舌苔检测、舌苔校准、舌特征提取和体质分类组成。此方法在个体层面上考虑了实例复杂性,以减少不同环境条件(如光照和分辨率)对图像分布不均的影响。复杂性感知分类方法是通用的,可以很容易地扩展到其他应用场景。

4 分类方法性能比较

本文主要以HANHUI在Kaggle网站上发布的有齿痕和无齿痕数据^[103]作为二分类的例子对比经典网络方法。Kaggle是2010年成立的预测模型和分析竞赛平台,目前被用作数据分析和机器学习的学习平台^[133]。数据包括1250张舌图像,其中546张有齿痕的异常图像和704张无齿痕的正常图像。本实验使用了5折交叉验证的方式,实验结果有以下6个指标评估:准确率(Acc)、查准率(P)、查全率(R)、F1度量(F1)、特异性(Spc)和虚警率(FAR)。表4列出了

表4 对于齿痕舌不同分类方法的评价指标

Table 4 Evaluation metrics for different tooth-marked tongue classification methods 单位:%

方法	Acc	P	R	F1	Spc	FAR
AlexNet ^[30]	59.76±4.17	36.84±57.74	9.17±9.53	14.38±16.36	99.14±0.00	0.86±0.00
VGG-11 ^[38]	66.27±0.81	92.38±11.28	25.51±3.22	39.75±2.80	98.00±3.38	2.00±3.38
VGG-13 ^[38]	92.37±0.46	94.35±5.03	88.08±5.22	90.99±0.62	95.71±4.35	4.29±4.35
VGG-16 ^[38]	88.51±5.24	91.93±3.26	80.92±12.45	85.74±7.99	94.43±2.58	5.57±2.58
VGG-19 ^[38]	68.44±2.96	84.99±10.61	36.70±17.56	49.00±12.74	93.14±8.42	6.86±8.42
GoogLeNet ^[39]	85.30±0.40	89.76±1.78	75.04±1.40	81.70±0.41	93.28±1.49	6.72±1.49
ResNet-18 ^[45]	75.10±1.98	78.28±3.83	59.82±1.91	67.65±2.24	87.00±2.86	13.00±2.86
ResNet-34 ^[45]	75.10±1.39	77.88±0.78	60.37±3.47	67.80±2.53	86.57±0.41	13.43±0.41
ResNet-50 ^[45]	74.38±0.41	79.36±4.90	56.51±6.51	65.75±3.20	88.29±4.36	11.71±4.36
ResNet-101 ^[45]	73.74±0.81	76.16±1.81	58.90±4.71	66.13±2.59	85.29±2.30	14.71±2.30
ResNet-152 ^[45]	71.73±0.46	74.25±2.89	54.50±5.11	62.71±2.56	85.15±3.38	14.85±3.38
MobileNet_v2 ^[134]	76.06±3.60	78.13±6.43	63.31±4.24	69.43±3.76	86.00±5.73	14.00±5.73
MobileNet_v3_small ^[135]	75.26±3.43	77.17±5.26	61.84±3.31	68.51±4.03	85.71±3.78	14.29±3.78
MobileNet_v3_large ^[135]	78.39±6.15	87.01±4.72	60.00±20.31	70.04±12.44	92.71±5.36	7.29±5.36
DenseNet121 ^[54]	65.30±1.86	100.00±0.00	20.74±4.24	34.25±5.96	100.00±0.00	0.00±0.00
DenseNet161 ^[54]	62.25±3.37	100.00±0.00	13.76±7.70	23.85±11.61	100.00±0.00	0.00±0.00
DenseNet169 ^[54]	65.06±1.63	99.00±2.89	20.37±3.48	33.71±4.89	99.86±0.41	0.14±0.41
DenseNet201 ^[54]	63.13±1.63	100.00±0.00	15.78±3.71	27.07±5.73	100.00±0.00	0.00±0.00
SE_ResNet-18 ^[57]	77.67±2.08	79.24±4.61	66.79±2.76	72.39±1.97	86.14±4.29	13.86±4.29
SE_ResNet-34 ^[57]	77.35±4.10	79.64±0.71	65.50±12.32	71.40±6.92	86.57±2.30	13.43±2.30
SE_ResNet-50 ^[57]	75.74±1.98	77.98±3.73	62.93±10.42	69.22±5.57	85.71±4.75	14.29±4.75
SE_ResNet-101 ^[57]	74.62±2.44	79.78±7.13	56.88±6.89	66.17±3.77	88.43±6.48	11.57±6.48
SE_ResNet-152 ^[57]	74.70±1.23	79.44±4.44	57.25±5.61	66.39±2.94	88.29±3.78	11.71±3.78

不同分类方法的各个评价指标的均值和方差。表5列出了经迁移学习后各个分类模型的评价指标的均值和方差。

AlexNet和DenseNet系列准确率较低。DenseNet系列的 Spc 接近于1,即被预测为负样本概率接近1。如果训练数据集有限,参数过多很容易产生过拟合,导致模型对有齿痕和无齿痕数据分布存在偏差,即DenseNet系列将样本直接分类为负样本。在AlexNet的基础上,VGG系列运用了更小的卷积核,并且加深了网络,达到了更好的效果。其中VGG-13对于齿痕舌的诊断采用5折交叉验证,其准确率均值为92.37%,方差为0.46%, $F1$ 、 Spc 和 FAR 分别为90.99%、95.71%和4.29%。无论是否使用迁移学习的

方法训练模型,VGG-13性能最好。GoogLeNet增加网络深度和宽度,但网络加深时,有时无法收敛,其准确率仅次于VGG-13和VGG-16。

SE_ResNet系列相比ResNet系列加入了通道注意力机制,即可自适应地校正通道特征。SE_ResNet系列增加少量参数,提升精度约2%。CNN的参数过多,实验往往达不到理想的效果。MobileNet系列是专注于移动端或嵌入式设备中的轻量级CNN网络,减少模型参数与运算量,结果表明在齿痕分类的精确度仍有提升的空间。而使用迁移学习的方法进行齿痕分类时,大多数模型性能都有所提高,在一定程度上克服了样本不足的局限性,提升了分类的效果。

表6对以上网络模型的优缺点进行了总结。数

表5 基于迁移学习的齿痕舌分类不同方法结果对比

Table 5 Evaluation metrics for different tooth-marked tongue classification methods

方法	based on transfer learning						单位 %
	Acc	P	R	F1	Spc	FAR	
AlexNet ^[30]	57.51±3.71	20.00±57.74	2.94±8.48	5.12±14.78	100.00±0.00	0.00±0.00	
VGG-11 ^[38]	64.02±0.24	72.93±20.42	23.12±14.83	33.00±10.49	95.86±11.96	4.14±11.96	
VGG-13 ^[38]	94.38±0.46	96.55±2.76	90.46±2.80	93.37±0.58	97.43±2.18	2.57±2.18	
VGG-16 ^[38]	93.98±0.23	94.20±1.73	91.93±2.31	93.03±0.39	95.57±1.48	4.43±1.48	
VGG-19 ^[38]	94.22±0.62	96.87±2.36	89.73±0.92	93.15±0.61	97.71±1.80	2.29±1.80	
GoogLeNet ^[39]	92.85±1.06	93.46±3.67	90.09±3.23	91.69±1.14	95.00±3.30	5.00±3.30	
ResNet-18 ^[45]	82.09±0.81	82.31±2.17	75.41±1.06	78.65±0.56	87.29±2.18	12.71±2.18	
ResNet-34 ^[45]	81.93±0.93	87.08±1.60	68.99±0.92	76.97±1.11	92.00±1.09	8.00±1.09	
ResNet-50 ^[45]	81.53±0.41	94.67±0.68	61.28±1.41	74.38±0.85	97.29±0.42	2.71±0.42	
ResNet-101 ^[45]	79.44±0.84	84.06±6.43	66.24±4.24	73.78±0.96	89.71±4.46	10.29±4.46	
ResNet-152 ^[45]	83.29±1.06	85.42±3.26	74.68±2.95	79.64±1.31	90.00±2.70	10.00±2.70	
MobileNet_v2 ^[134]	81.85±0.93	87.06±0.97	68.81±2.31	76.83±1.51	92.00±0.72	8.00±0.72	
MobileNet_v3_small ^[135]	77.75±1.23	83.42±1.01	61.65±2.42	70.76±1.87	90.29±0.41	9.71±0.41	
MobileNet_v3_large ^[135]	81.93±0.40	86.57±0.95	69.54±0.92	77.10±0.55	91.57±0.72	8.43±0.72	
DenseNet121 ^[54]	56.22±0.00	0.00±0.00	0.00±0.00	0.00±0.00	100.00±0.00	0.00±0.00	
DenseNet161 ^[54]	56.22±0.00	0.00±0.00	0.00±0.00	0.00±0.00	100.00±0.00	0.00±0.00	
DenseNet169 ^[54]	56.22±0.00	0.00±0.00	0.00±0.00	0.00±0.00	100.00±0.00	0.00±0.00	
DenseNet201 ^[54]	56.22±0.00	0.00±0.00	0.00±0.00	0.00±0.00	100.00±0.00	0.00±0.00	

表6 神经网络模型的优缺点

Table 6 Advantages and disadvantages of neural network models

方法	优点	缺点
AlexNet ^[30]	训练速度快,提出系列防止过拟合方法	计算量大
VGG ^[38]	堆叠多个小卷积核,网络更深,效果更好	训练时间长
GoogLeNet ^[39]	提出并联结构,增加网络深度和宽度,减少参数	网络加深时,有时无法收敛
ResNet ^[45]	残差学习解决退化问题,使用BN加速训练	存在冗余
SENet ^[57]	实现方便,计算量小,即插即用	模型参数和运行时间有所增加
MobileNet ^[134-135]	减少模型参数与运算量	准确率有待提升
DenseNet ^[54]	减轻梯度消失,增强特征传播,加强特征复用,减少权重参数	对于小样本数据集容易过拟合

据集齿痕舌体的边缘信息目标较小,凹凸程度不同,并且特征相对不明显。如果数据量不够,在训练模型时容易出现难收敛或者过拟合的问题,导致模型分类性能变差。因此,需要收集更多的高质量数据以获得有临床意义的诊断结果。

5 存在的问题和发展趋势

虽然深度学习在自动提取高维特征方面比传统方法更具优势,但是仍然存在一些挑战和局限性。本文将从以下四方面对深度学习在舌象分类中存在的问题和发展趋势进行讨论。

(1)构建数据集。大多研究人员自建非公开的舌诊图像数据集,因而无法在相同的数据集上对分类模型进行性能比较。而深度学习网络需要海量和高质量数据,训练成本高,使用较小数据规模的网络分类是未来发展的趋势。舌诊图像数据集一般通过普通的摄像机、照相机或中医舌诊仪采集,由于使用的设备或舌诊图像采集系统不同,并且没有任何的通用标准,导致所获得的舌诊图像质量有显著差异。随着互联网与云计算不断发展,移动设备成为采集舌诊图像数据的一种工具,但数据质量有限,它需要更大的数据集和更好的模型才能达到较好的效果。此外,被采集者无法将舌体完整伸展导致脏腑体表区域不完善,饮食或药物可能引起舌苔颜色异常或舌体异常,颤抖造成图像模糊。因此舌诊图像的质量评判尤为重要,舌诊图像质量的好坏影响最终的分类结果。

(2)开放环境。基于早期神经网络、卷积神经网络、区域卷积神经网络、迁移学习以及其他方法在封闭环境下取得了较好的鲁棒性。然而,提高中医辅助诊断在真实诊疗环境下的准确率以及受试者的体验舒适度是一发展趋势。虽然这种开放环境不易控制环境复杂性和动态性,如光线、温度等,但可使被采集者有较为自然的生理心理状态,更易于还原真实情况和真实的中医诊断环境,从而克服现有大多数研究采用暗箱式数码相机拍照的封闭形式,存在的易引发被采集者产生心理压力而不能保持血气平和与舌尖自然下垂的问题。因此,从自然的中医诊断环境出发,搭建开放舌诊视频图像采集环境,降低对被采集者的约束程度,为一次性获得较多有效图像,采取视频采集,并从中筛选有效图像以及动态特征,得到视频图像数据集。

(3)提取局部和全局特征。近年来,研究者将局

部及全局特征提取^[101]用于舌象诊断,从临床数据中学习有用的特征,并利用获得的特征采用有监督的技术将病人的健康状况进行划分。常见的舌象全局特征包含舌体的神、色、形、态。全局特征具有不变性且计算简单,但其有特征维数高、计算量大等缺点。常见的舌象局部特征包含舌形、舌苔、舌质、纹理、动态及舌下络脉等。局部特征是从图像区域中抽取的特征,其特征间的相关性小,特征描述维度低,易于实现快速匹配。注意力机制能够提高CNN的可解释性和发现新的舌特征(例如与疾病相关的区域中的特定颜色)以及提取局部和全局特征的能力,可以对小样本的舌象分类提供重要的帮助。未来CNN框架的设计可以在不同层次引入注意力机制,使计算机视觉更接近人类视觉能力。

(4)单标签与多标签分类。目前,大多数CNN模型仅支持单标签数据^[89],将任务简化为单标签分类问题。但是同一样本的舌象特征复杂且相互有关联,临床患者较少有单一证型,往往出现多个证型兼夹的复杂情况,属于典型的多标签学习问题。此外,舌诊病例样本具有不均衡特点,导致难以直接训练基于深度学习的多标签分类模型。针对这些问题,训练多个单标签网络,从而将多分类问题转换为二分类问题,以此来消除多标签的歧义性;舌象特征相互有关联,分别进行特征提取和分类,忽略了它们之间存在关联的有效信息。此外,Transformer与CNN不一样,它没有卷积、Pooling等操作,也没有循环,然而在中医图像单标签与多标签分类任务中应用较少,后续可以广泛地应用。

6 结束语

深度学习在医学领域的发展依赖于医学大数据的积累。在疾病治疗方面,深度学习不仅可以检测到病灶区域,还可以对特定病灶进行判别和分类。如何将深度学习更好地应用于医疗的各个阶段,成为一项更具挑战性的任务。本文首先对深度学习的舌象分类现状进行调查、研究和总结,以便为该领域的研究人员提供参考,最后对未来的研究方向进行探讨和分析。随着深度学习技术的快速发展和计算能力的提高,深度学习在舌象分类领域将得到了更广泛的应用。

参考文献:

- [1] MACIOCIA G. Tongue diagnosis in Chinese medicine

- [M]. Seattle: Eastland Press, 1995.
- [2] ZHANG D. Automated biometrics: technologies and systems [M]. Berlin, Heidelberg: Springer, 2000.
- [3] BARBARA K. Atlas of Chinese of tongue diagnosis[M]. Seattle: Eastland Press, 1998.
- [4] PANG B, ZHANG D, LI N, et al. Computerized tongue diagnosis based on Bayesian networks[J]. IEEE Transactions on Biomedical Engineering, 2004, 51(10): 1803-1810.
- [5] ZHI L, ZHANG D, YAN J Q, et al. Classification of hyperspectral medical tongue images for tongue diagnosis[J]. Computerized Medical Imaging and Graphics, 2007, 31(8): 672-678.
- [6] QI Z, TU L P, CHEN J B, et al. The classification of tongue colors with standardized acquisition and ICC profile correction in traditional Chinese medicine[J]. BioMed Research International, 2016: 3510807.
- [7] 王昇, 刘升华, 王丽婷. 舌诊图像点刺和瘀点的识别与提取[J]. 计算机工程与科学, 2017, 39(6): 1126-1132.
WANG S, LIU K H, WANG L T. Tongue spots and petechiae recognition and extraction in tongue diagnosis images[J]. Computer Engineering & Science, 2017, 39(6): 1126-1132.
- [8] YAMAMOTO S, TSUMURA N, NAKAGUCHI T, et al. Principal component vector rotation of the tongue color spectrum to predict "Mibyou" (disease-oriented state) [J]. International Journal of Computer Assisted Radiology and Surgery, 2011, 6(2): 209-215.
- [9] 刘畅. 基于决策树及神经网络对儿童抽动障碍肾虚不足证舌象的研究[D]. 济南: 山东中医药大学, 2020.
LIU C. A study of tongue feature in children with tic disorders of kidney emotion deficiency based on decision tree and neural network[D]. Jinan: Shandong University of Traditional Chinese Medicine, 2020.
- [10] 王奕然, 张新峰. 基于 AdaBoost 级联框架的舌色分类[J]. 北京生物医学工程, 2020, 39(1): 8-14.
WANG Y R, ZHANG X F. Tongue color classification based on AdaBoost cascade framework[J]. Beijing Biomedical Engineering, 2020, 39(1): 8-14.
- [11] BHARATHI M, PRASAD D, VENKATAKRISHNAMOORTHY T, et al. Diabetes diagnostic method based on tongue image classification using machine learning algorithms[J]. Journal of Pharmaceutical Negative Results, 2022, 13(4): 1247-1250.
- [12] GORUR K, BOZKURT M R, BASCIL M S, et al. GKP signal processing using deep CNN and SVM for tongue-machine interface[J]. Traitement du Signal, 2019, 36(4): 319-329.
- [13] OBAFEMI-AJAYI T, KANAWONG R, XU D, et al. Features for automated tongue image shape classification[C]// Proceedings of the 2012 IEEE International Conference on Bioinformatics and Biomedicine Workshops, Philadelphia, Oct 4-7, 2012. Washington: IEEE Computer Society, 2012: 273-279.
- [14] 李宗润. 基于深度学习技术的舌体分割模型研究与舌象智能化应用探索[D]. 成都: 成都中医药大学, 2020.
LI Z R. The study of tongue segmentation model based on deep learning technology and application exploration of the intelligence of tongue image[D]. Chengdu: Chengdu University of Traditional Chinese Medicine, 2020.
- [15] 李红岩, 李灿, 郎许锋, 等. 中医四诊智能化研究现状及热点分析[J]. 南京中医药大学学报, 2022, 38(2): 180-186.
LI H Y, LI C, LANG X F, et al. Analysis of the research status and hot spot of intelligent four-diagnosis in TCM[J]. Journal of Nanjing University of Traditional Chinese Medicine, 2022, 38(2): 180-186.
- [16] 牛富泉. 基于深度学习的中医舌像分类模型研究[D]. 成都: 电子科技大学, 2021.
NIU F Q. Research on tongue image classification in traditional Chinese medicine based on deep learning[D]. Chengdu: University of Electronic Science and Technology of China, 2021.
- [17] 鲍康. 基于舌头检测与舌象识别的中医智能诊断系统的设计与实现[D]. 北京: 北京交通大学, 2021.
BAO K. Design and implementation of TCM intelligent diagnosis system based on tongue detection and tongue manifestation recognition[D]. Beijing: Beijing Jiaotong University, 2021.
- [18] 刘飞, 张俊然, 杨豪. 基于深度学习的糖尿病患者的分类识别 [J]. 计算机应用, 2018, 38(S1): 39-43.
LIU F, ZHANG J R, YANG H. Classification and recognition of diabetes mellitus based on deep learning[J]. Journal of Computer Application, 2018, 38(S1): 39-43.
- [19] SUN Y, DAI S, LI J, et al. Tooth-marked tongue recognition using gradient-weighted class activation maps[J]. Future Internet, 2019, 11(2): 45.
- [20] 王东军, 朱青青, 余静寅, 等. 基于CiteSpace的舌诊客观化研究领域科学知识图谱与可视化分析[J]. 世界科学技术: 中医药现代化, 2021, 23(1): 283-291.
WANG D J, LI Q Q, YU J Y, et al. Atlas and visualization analysis of scientific knowledge in the field of tongue objectification research based on CiteSpace[J]. World Science and Technology: Modernization of Traditional Chinese Medicine, 2021, 23(1): 283-291.
- [21] 李渊彤, 罗裕升, 朱珍民. 基于深度学习的舌象特征分析 [J]. 计算机科学, 2020, 47(11): 148-158.
LI Y T, LUO Y S, ZHU Z M. Tongue image analysis in

- traditional Chinese medicine based on deep learning[J]. Computer Science, 2020, 47(11): 148-158.
- [22] MCCULLOCH W S, PITTS W. A logical calculus of the ideas immanent in nervous activity[J]. The Bulletin of Mathematical Biophysics, 1943, 5(4): 115-133.
- [23] CHEN L, LI S, BAI Q, et al. Review of image classification algorithms based on convolutional neural networks[J]. Remote Sensing, 2021, 13(22): 4712.
- [24] ROSENBLATT F. The perceptron: a probabilistic model for information storage and organization in the brain[J]. Psychological Review, 1958, 65(6): 386.
- [25] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [26] 邢甜甜. 基于卷积神经网络的舌象模式分类研究[D]. 秦皇岛: 燕山大学, 2018.
- XING T T. Classification of tongue pattern based on convolution neural network[D]. Qinhuangdao: Yanshan University, 2018.
- [27] ZHOU L, PAN S, WANG J, et al. Machine learning on big data: opportunities and challenges[J]. Neurocomputing, 2017, 237: 350-361.
- [28] MANGASARIAN O L, MUSICANT D R. Data discrimination via nonlinear generalized support vector machines[M]. Berlin, Heidelberg: Springer, 2001.
- [29] 张珂, 冯晓晗, 郭玉荣, 等. 图像分类的深度卷积神经网络模型综述[J]. 中国图象图形学报, 2021, 26(10): 2305-2325.
- ZHANG K, FENG X H, GUO Y R, et al. Overview of deep convolutional neural networks for image classification[J]. Journal of Image and Graphics, 2021, 26(10): 2305-2325.
- [30] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [31] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]//LNCS 8689: Proceedings of the 13th European Conference on Computer Vision, Zurich, Sep 6-12, 2014. Cham: Springer, 2014: 818-833.
- [32] XUE Y, LI X, CUI Q, et al. Cracked tongue recognition based on deep features and multiple- instance SVM[C]//LNCS 11165: Proceedings of the 19th Pacific-Rim Conference on Multimedia, Hefei, Sep 21-22, 2018. Cham: Springer, 2018: 642-652.
- [33] HUO C M, ZHENG H, SU H Y, et al. Tongue shape classification integrating image preprocessing and convolution neural network[C]//Proceedings of the 2017 2nd Asia-Pacific Conference on Intelligent Robot Systems, Wuhan, Jun 16-18, 2017. Piscataway: IEEE, 2017: 42-46.
- [34] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]//Proceedings of the 32nd International Conference on Machine Learning, Lille, Jul 6-11, 2015: 448-456.
- [35] 肖庆新, 张菁, 张辉, 等. 基于轻型卷积神经网络的舌苔颜色分类方法[J]. 测控技术, 2019, 38(3): 26-31.
- XIAO Q X, ZHANG J, ZHANG H, et al. Tongue coating color classification based on shallow convolutional neural network[J]. Measurement & Control Technology, 2019, 38 (3): 26-31.
- [36] JIA Y Q, SHELHAMER E, DONAHUE J, et al. Caffe: convolutional architecture for fast feature embedding[C]//Proceedings of the 2014 ACM International Conference on Multimedia, Orlando, Nov 3-7, 2014. New York: ACM, 2014: 675-678.
- [37] 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, Beijing, Mar 10-12, 2017. Piscataway: IEEE, 2017: 725-729.
- [38] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv:1409.1556, 2014.
- [39] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern recognition, Boston, Jun 7-12, 2015. Washington: IEEE Computer Society, 2015: 1-9.
- [40] RORA S, BHASKARA A, GE R, et al. Provable bounds for learning some deep representations[C]//Proceedings of the 31st International Conference on Machine Learning, Beijing, Jun 21-26, 2014: 584-592.
- [41] LIN M, CHEN Q, YAN S. Network in network[J]. arXiv: 1312.4400, 2013.
- [42] FU S, ZHENG H, YANG Z, et al. Computerized tongue coating nature diagnosis using convolutional neural network [C]//Proceedings of the 2017 IEEE 2nd International Conference on Big Data Analysis, Beijing, Mar 10-12, 2017. Piscataway: IEEE, 2017: 730-734.
- [43] OTSU N. A threshold selection method from gray-level histograms[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1): 62-66.
- [44] SMITH A R. Color gamut transform pairs[J]. ACM SIGGRAPH Computer Graphics, 1978, 12(3): 12-19.
- [45] HE K M, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Jun 27-30, 2016. Washington: IEEE Computer Society, 2016: 770-778.
- [46] LI J, ZHANG Z, ZHU X, et al. Automatic classification framework of tongue feature based on convolutional neural networks[J]. Micromachines, 2022, 13(4): 501.

- [47] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]// LNCS 9351: Proceedings of the 2015 International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Oct 5-9, 2015. Cham: Springer, 2015: 234-241.
- [48] 邵尤伟. 基于深度学习的智能舌诊方法研究[D]. 厦门: 厦门大学, 2018.
- SHAO Y W. Research on intelligent tongue diagnosis based on deep learning[D]. Xiamen: Xiamen University, 2018.
- [49] ZHANG D, ZHANG H Z, ZHANG B. Statistical analysis of tongue color and its applications in diagnosis[M]. Berlin, Heidelberg: Springer, 2017.
- [50] CHANG W H, CHU H T, CHANG H H. Tongue fissure visualization with deep learning[C]//Proceedings of the 2018 Conference on Technologies and Applications of Artificial Intelligence, Taichung, China, Nov 30- Dec 2, 2018. Washington: IEEE Computer Society, 2018: 14-17.
- [51] SELVARAJU R R, 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, Venice, Oct 22-29, 2017. Washington: IEEE Computer Society, 2017: 618-626.
- [52] TANG Y, SUN Y, CHIANG J Y, et al. Research on multiple-instance learning for tongue coating classification[J]. IEEE Access, 2021, 9: 66361-66370.
- [53] ANDREWS S, TSOCHANTARIDIS I, HOFMANN T. Support vector machines for multiple-instance learning[C]//Proceedings of the 2002 Conference on Neural Information Processing Systems, Vancouver, Dec 9- 14, 2002. Cambridge: MIT Press, 2002: 561-568.
- [54] 孙萌, 张新峰. 基于TripletLoss损失函数的舌象分类方法研究[J]. 北京生物医学工程, 2020, 39(2): 131-137.
- SUN M, ZHANG X F. Tongue image classification based TripletLoss metric[J]. Beijing Biomedical Engineering, 2020, 39(2): 131-137.
- [55] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Jul 21-26, 2017. Washington: IEEE Computer Society, 2017: 2261-2269.
- [56] 陈慧贞. 基于卷积神经网络的舌象辨识模型应用研究[D]. 秦皇岛: 燕山大学, 2019.
- CHEN H Z. Research on application of tongue recognition model based on convolutional neural network[D]. Qinhuangdao: Yanshan University, 2019.
- [57] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Piscataway: IEEE, 2018: 7132-7141.
- [58] 骆明楠. 面向智能舌诊的注意力卷积网络研究[D]. 广州: 华南理工大学, 2020.
- LUO M N. Research on attention convolution networks for intelligent tongue diagnosis[D]. Guangzhou: South China University of Technology, 2020.
- [59] GIRSHICK R, DONAHUE J, DARELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, Jun 23-28, 2014. Washington: IEEE Computer Society, 2014: 580-587.
- [60] AZIZ L, SALAM M S B H, SHEIKH U, et al. Exploring deep learning-based architecture, strategies, applications and current trends in generic object detection: a comprehensive review[J]. IEEE Access, 2020, 8: 170461-170495.
- [61] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C]//Proceedings of the Annual Conference on Neural Information Processing Systems 2015, Montreal, Dec 7-12, 2015: 91-99.
- [62] GAVRILESCU R, ZET C, FOŞALĂU C, et al. Faster R-CNN: an approach to real-time object detection[C]//Proceedings of the 2018 International Conference and Exposition on Electrical and Power Engineering, Iasi, Oct 18-19, 2018. Piscataway: IEEE, 2018: 165-168.
- [63] SHAO Q, LI X, FU Z. Recognition of teeth-marked tongue based on gradient of concave region[C]//Proceedings of the 2014 International Conference on Audio, Language and Image Processing, Shanghai, Jul 7-9, 2014. Piscataway: IEEE, 2014: 968-972.
- [64] WANG H, ZHANG X, CAI Y. Research on teeth marks recognition in tongue image[C]//Proceedings of the 2014 International Conference on Medical Biometrics, Shenzhen, May 30-Jun 1, 2014. Piscataway: IEEE, 2014: 80-84.
- [65] LI X, ZHANG Y, CUI Q, et al. Tooth-marked tongue recognition using multiple instance learning and CNN features[J]. IEEE Transactions on Cybernetics, 2018, 49(2): 380-387.
- [66] GHOLAMI E, TABBAKH S R K, KHEIRABADI M. Proposing method to increase the detection accuracy of stomach cancer based on colour and lint features of tongue using CNN and SVM[J]. arXiv:2011.09962, 2020.
- [67] 杜春慧. 中医舌质特征的机器学习模型研究[D]. 成都: 电

- 子科技大学, 2020.
- DU C H. Research on machine learning model of tongue characteristics in traditional Chinese medicine[D]. Chengdu: University of Electronic Science and Technology of China, 2020.
- [68] CHEN P, MEN S, LIN H, et al. Detection of local lesions in tongue recognition based on improved Faster R-CNN[C]// Proceedings of the 2021 6th International Conference on Computational Intelligence and Applications, Xiamen, Jun 11-13, 2021. Piscataway: IEEE, 2021: 165-168.
- [69] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]// Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Oct 22-29, 2017. Washington: IEEE Computer Society, 2017: 2980-2988.
- [70] LI J, CHEN Q, HU X, et al. Establishment of noninvasive diabetes risk prediction model based on tongue features and machine learning techniques[J]. International Journal of Medical Informatics, 2021, 149: 104429.
- [71] PENG C D, WANG L, JIANG D M, et al. Establishing and validating a spotted tongue recognition and extraction model based on multiscale convolutional neural network[J]. Digital Chinese Medicine, 2022, 5(1): 49-58.
- [72] 刘佳丽, 孙自强. 基于Double-D算法的舌像检测[J]. 计算机工程与设计, 2020, 41(7): 2025-2030.
LIU J L, SUN Z Q. Tongue image detection based on Double-D algorithm[J]. Computer Engineering and Design, 2020, 41(7): 2025-2030.
- [73] YEN S J, YE S C, LEE C L. Traditional Chinese medicine online intelligent tongue diagnosis bot[C]//Proceedings of the 2021 International Symposium on Intelligent Signal Processing and Communication Systems, Hualien, China, Nov 16-19, 2021. Piscataway: IEEE, 2021: 1-2.
- [74] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[J]. arXiv: 2004.10934, 2020.
- [75] 颜建军, 李东旭, 郭睿, 等. 基于深度学习和随机森林的齿痕舌分类研究[J]. 中华中医药学刊, 2022, 40(2): 19-22.
YAN J J, LI D X, GUO R, et al. Research on classification of dentate tongue based on deep learning and random forest [J]. Chinese Archives of Traditional Chinese Medicine, 2022, 40(2): 19-22.
- [76] WENG H, LI L, LEI H, et al. A weakly supervised tooth-mark and crack detection method in tongue image[J]. Concurrency and Computation: Practice and Experience, 2021, 33(16): e6262.
- [77] 高爽, 徐巧枝. 迁移学习方法在医学图像领域的应用综述 [J]. 计算机工程与应用, 2021, 57(24): 39-50.
GAO S, XU Q Z. Review of application of transfer learning in medical image field[J]. Computer Engineering and Applications, 2021, 57(24): 39-50.
- [78] CHEN Z, ZHANG X, HUANG W. The similar sparse domain adaptation illustrated by the case of TCM tongue inspection[C]//Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine, Seoul, Dec 16-19, 2020. Piscataway: IEEE, 2020: 1520-1525.
- [79] 刘婧玮. 基于卷积神经网络的中医舌象辨识人工智能方法学研究[D]. 北京: 北京中医药大学, 2020.
- LIU J W. Artificial intelligence methodology for Chinese medicine tongue recognition based on convolutional neural network[D]. Beijing: Beijing University of Chinese Medicine, 2020.
- [80] 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.
- [81] 邱童. 基于深度学习与多特征融合的舌象诊断算法[J]. 现代信息科技, 2020, 4(1): 63-65.
QIU T. Tongue image diagnosis algorithm based on deep learning and multi-feature fusion[J]. Modern Information Technology, 2020, 4(1): 63-65.
- [82] HU Y, WEN G, LIAO H, et al. Automatic construction of Chinese herbal prescriptions from tongue images using CNNs and auxiliary latent therapy topics[J]. IEEE Transactions on Cybernetics, 2019, 51(2): 708-721.
- [83] SONG C, WANG B, XU J. Classifying tongue images using deep transfer learning[C]//Proceedings of the 2020 5th International Conference on Computational Intelligence and Applications, Beijing, Jun 19-21, 2020. Piscataway: IEEE, 2020: 103-107.
- [84] SADASIVAN S, SIVAKUMAR T T, JOSEPH A P, et al. Tongue print identification using deep CNN for forensic analysis[J]. Journal of Intelligent & Fuzzy Systems, 2020, 38(5): 6415-6422.
- [85] 杨晶东, 张朋. 基于迁移学习的全连接神经网络舌象分类方法[J]. 第二军医大学学报, 2018, 39(8): 897-902.
YANG J D, ZHANG P. Tongue image classification method based on transfer learning and fully connected neural network[J]. Academic Journal of Naval Medical University, 2018, 39(8): 897-902.
- [86] 翟鹏博, 杨浩, 宋婷婷, 等. 融合注意力机制的多阶段舌象分类算法[J]. 计算机工程与设计, 2021, 42(6): 1606-1613.
ZHAI P B, YANG H, SONG T T, et al. Multi-stage image classification algorithm incorporating attention mechanism [J]. Computer Engineering and Design, 2021, 42(6): 1606-1613.

- [87] KONG X, RUI Y, DONG X, et al. Tooth-marked tongue recognition based on mask scoring R-CNN[C]//Proceedings of the 2020 2nd International Conference on Artificial Intelligence Technologies and Application, Dalian, Aug 21-23, 2020.
- [88] 刘梦, 王曦廷, 周璐, 等. 基于深度学习与迁移学习的中医舌象提取识别研究[J]. 中医杂志, 2019(10): 835-840.
- LIU M, WANG X Y, ZHOU L, et al. Study on extraction and recognition of traditional Chinese tongue manifestation: based on deep learning and migration learning[J]. Journal of Traditional Chinese Medicine, 2019(10): 835-840.
- [89] 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 Bio-Medical and Health Informatics, 2020, 24(9): 2481-2489.
- [90] 汤一平, 王丽冉, 何霞, 等. 基于多任务卷积神经网络的舌象分类研究[J]. 计算机科学, 2018, 45(12): 255-261.
- TANG Y P, WANG L R, HE X, et al. Classification of tongue image based on multi-task deep convolution neural network[J]. Computer Science, 2018, 45(12): 255-261.
- [91] 王爱民, 赵忠旭, 沈兰荪. 中医舌象自动分析中舌色、苔色分类方法的研究[J]. 北京生物医学工程, 2000, 19(3): 136-141.
- WANG A M, ZHAO Z X, SHEN L S. Research on the tongue color classification in automatic analysis of traditional Chinese medicine[J]. Beijing Biomedical Engineering, 2000, 19(3): 136-141.
- [92] MIKA S, RATSCH G, WESTON J, et al. Fisher discriminant analysis with kernels[C]//Proceedings of the 1999 IEEE Workshop on Neural Networks for Signal Processing, Madison, Aug 25, 1999. Piscataway: IEEE, 1999: 41-48.
- [93] 张思远, 翟宏琛, 梁艳梅, 等. 模糊相关中的加权算法及其在彩色图像检索中的应用[J]. 中国科学: G辑物理学、力学、天文学, 2004, 34(1): 60-68.
- ZHANG S Y, ZHAI H C, LIANG Y M, et al. Weighting algorithm in fuzzy correlation and its application in color image retrieval[J]. Scientia Sinica: Physica, Mechanica & Astronomica, 2004, 34(1): 60-68.
- [94] QIN J, MEN G J, WU G Z, et al. The retrieval of the medical tongue-coating images using fuzzy CMAC neural network[C]//Proceedings of the 2005 IEEE International Conference on Industrial Technology, Hong Kong, China, Dec 14-17, 2005. Piscataway: IEEE, 2005: 465-468.
- [95] QIU T. Tongue identification for small samples based on meta learning[C]//Proceedings of the 2020 International Conference on Computer Information and Big Data Applications, Guiyang, Apr 17-19, 2020. Piscataway: IEEE, 2020: 295-299.
- [96] XIAO Q, ZHANG H, ZHANG J, et al. Texture analysis of tongue coating in traditional Chinese medicine based on transfer learning and multi-model decision[J]. Sensing and Imaging, 2021, 22(1): 1-13.
- [97] HU Y, WEN G, LUO M, et al. Fully-channel regional attention network for disease-location recognition with tongue images[J]. Artificial Intelligence in Medicine, 2021, 118: 102110.
- [98] XU Y, WEN G, YANG P, et al. Task-coupling elastic learning for physical sign-based medical image classification [J]. IEEE Journal of Biomedical and Health Informatics, 2021, 26(2): 626-637.
- [99] DAI Y, WANG G. Analyzing tongue images using a conceptual alignment deep autoencoder[J]. IEEE Access, 2018, 6: 5962-5972.
- [100] MENG D, CAO G, DUAN Y, et al. Tongue images classification based on constrained high dispersal network [J]. Evidence-Based Complementary and Alternative Medicine, 2017(3): 7452427.
- [101] 卓力, 孙亮亮, 张辉, 等. 有噪声标注情况下的中医舌色分类方法[J]. 电子与信息学报, 2022, 44(1): 89-98.
- ZHUO L, SUN L L, ZHANG H, et al. TCM tongue color classification method under noisy labeling[J]. Journal of Electronics & Information Technology, 2022, 44(1): 89-98.
- [102] RAJAKUMARAN S, SASIKALA J. Improvement in tongue color image analysis for disease identification using deep learning based depthwise separable convolution model[J]. Indian Journal of Computer Science and Engineering, 2021, 12: 21-34.
- [103] HANHUI. Tooth-marked-tongue[EB/OL]. (2020-03-10)[2022-07-25]. <http://www.kaggle.com/datasets/clearhanhui/byiesheji>.
- [104] BioHit. TongueImageDataset[EB/OL]. (2014-09-05) [2022-07-25]. <http://github.com/BioHit/TongueImageDatabase>.
- [105] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: synthetic minority over-sampling technique[J]. Journal of Artificial Intelligence Research, 2002, 16: 321-357.
- [106] CHEN Z, ZHANG X, QIU R. Application of artificial intelligence in tongue diagnosis of traditional Chinese medicine: a review[J]. TMR Modern Herbal Medicine, 2021, 4(2): 14-30.
- [107] 刘国正. 卷积神经网络在中医舌象分类模型中的应用研究[D]. 长春: 吉林大学, 2018.
- LIU G Z. Research on application of traditional Chinese medicine tongue images classification based on CNN[D]. Changchun: Jilin University, 2018.
- [108] CHIU C C. A novel approach based on computerized image analysis for traditional Chinese medical diagnosis of the

- tongue[J]. Computer Methods and Programs in Biomedicine, 2000, 61(2): 77-89.
- [109] GUI M, ZHANG X, HU G, et al. A study on tongue image color description based on label distribution learning[C]// Proceedings of the 8th International Conference on Biomedical Engineering and Informatics, Shenyang, Oct 14-16, 2015. Piscataway: IEEE, 2015: 148-152.
- [110] ZHANG H, WANG K, ZHANG D, et al. Computer aided tongue diagnosis system[C]//Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, Shanghai, Jan 17-18, 2005. Piscataway: IEEE, 2006: 6754-6757.
- [111] HUI S C, HE Y, THACH D T C. Machine learning for tongue diagnosis[C]//Proceedings of the 2007 6th International Conference on Information, Communications & Signal Processing, Singapore, Dec 10-13, 2007. Piscataway: IEEE, 2007: 1-5.
- [112] HU J W, YAN Z Z, JIANG J H. Classification of fissured tongue images using deep neural networks[J]. Technology and Health Care, 2022, 30(S1): 271-283.
- [113] HSU P C, WU H K, HUANG Y C, et al. The tongue features associated with type 2 diabetes mellitus[J]. Medicine, 2019, 98(19): e15567.
- [114] ZHANG B, KUMAR B V, ZHANG D. Detecting diabetes mellitus and nonproliferative diabetic retinopathy using tongue color, texture, and geometry features[J]. IEEE Transactions on Biomedical Engineering, 2013, 61(2): 491-501.
- [115] LI J, HU X, TU L, et al. Diabetes tongue image classification using machine learning and deep learning[R]. Shanghai University of Traditional Chinese Medicine, 2021: 1-20.
- [116] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[J]. arXiv:201011929, 2020.
- [117] VASWANI A, SHAZEE N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the Annual Conference on Neural Information Processing Systems 2017, Long Beach, Dec 4-9, 2017. Red Hook: Curran Associates, 2017: 5998-6008.
- [118] PRAJAPATI M D, NAMDEV N. Prediction of heart disease risk based on deep learning[J]. International Research Journal of Modernization in Engineering Technology and Science, 2022, 4: 1011-1016.
- [119] SU J, LI Z, HUANG M, et al. Triglyceride glucose index for the detection of the severity of coronary artery disease in different glucose metabolic states in patients with coronary heart disease: a RCSCD-TCM study in China[J]. Cardiovascular Diabetology, 2022, 21(1): 1-9.
- [120] 叶桦, 冯全生, 严小英, 等. 基于人工神经网络的糖尿病合并冠心病舌脉象征型预测研究[J]. 中华中医药杂志, 2020, 35(10): 5184-5187.
- YE H, FENG Q S, YAN X Y, et al. Study of tongue and pulses syndrome prediction of diabetes mellitus combined with coronary heart disease based on artificial neural network [J]. China Journal of Traditional Chinese Medicine and Pharmacy, 2020, 35(10): 5184-5187.
- [121] LIANG K, HUANG X, CHEN H, et al. Tongue diagnosis and treatment in traditional Chinese medicine for severe COVID-19: a case report[J]. Annals of Palliative Medicine, 2020, 9(4): 2400-2407.
- [122] WANG X, WANG X, LOU Y, et al. Constructing tongue coating recognition model using deep transfer learning to assist syndrome diagnosis and its potential in non-invasive ethnopharmacological evaluation[J]. Journal of Ethnopharmacology, 2022, 285: 114905.
- [123] 中华中医药学会. 中医体质分类与判定: ZYYXH/T 157—2009[S]. 2009.
- China Association of Chinese Medicine. Classification and determination of constitution in TCM: ZYYXH/T 157—2009[S]. 2009.
- [124] ZHOU H, HU G, ZHANG X. Constitution identification of tongue image based on CNN[C]//Proceedings of the 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, Beijing, Oct 13-15, 2018. Piscataway: IEEE, 2018: 1-5.
- [125] 周浩, 胡广芹, 张新峰. 基于舌图像深度特征融合的中医体质分类方法研究[J]. 北京生物医学工程, 2020, 39(3): 221-226.
- ZHOU H, HU G Q, ZHANG X F. Preliminary study on traditional Chinese medicine constitution classification method based on tongue image feature fusion[J]. Beijing Biomedical Engineering, 2020, 39(3): 221-226.
- [126] 谢佳澄. 深度学习在中医舌象分类中的应用与实践[D]. 长春: 吉林大学, 2019.
- XIE J C. Application and practice of deep learning in classification of tongue image in traditional Chinese medicine [D]. Changchun: Jilin University, 2019.
- [127] LI H, WEN G, ZENG H. Natural tongue physique identification using hybrid deep learning methods[J]. Multimedia Tools and Applications, 2019, 78(6): 6847-6868.
- [128] LI T, WU C, MA Y. Multi-label constitution identification based on tongue image in traditional Chinese medicine [C]//Proceedings of the 2021 China Automation Congress, Beijing, Oct 22-24, 2021. Piscataway: IEEE, 2021: 1617-1622.
- [129] JIANG T, HU X J, YAO X H, et al. Tongue image quality

- assessment based on a deep convolutional neural network [J]. BMC Medical Informatics and Decision Making, 2021, 21(1): 1-14.
- [130] BARDOU D, ZHANG K, AHMAD S M. Lung sounds classification using convolutional neural networks[J]. Artificial Intelligence in Medicine, 2018, 88: 58-69.
- [131] GANDOMKAR Z, BRENNAN P C, MELLO-THOMS C. MuDeRN: multi-category classification of breast histopathological image using deep residual networks[J]. Artificial Intelligence in Medicine, 2018, 88: 14-24.
- [132] MA J, WEN G, WANG C, et al. Complexity perception classification method for tongue constitution recognition [J]. Artificial Intelligence in Medicine, 2019, 96: 123-133.
- [133] JIANG Y, CHI Z. A CNN model for semantic person part segmentation with capacity optimization[J]. IEEE Transactions on Image Processing, 2018, 28(5): 2465-2478.
- [134] SANDLER M, HOWARD A, ZHU M, et al. MobileNetv2: inverted residuals and linear bottle-necks[C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-23, 2018. Piscataway: IEEE, 2018: 4510-4520.
- [135] HOWARD A, SANDLER M, CHU G, et al. Searching for MobileNetv3[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Oct 27-Nov 2, 2019. Piscataway: IEEE, 2019: 1314-1324.



吴欣(1997—),女,重庆人,硕士研究生,主要研究方向为计算机视觉、图像处理。
WU Xin, born in 1997, M.S. candidate. Her research interests include computer vision and image processing.



徐红(1967—),女,北京人,博士,教授,研究方向为智能医疗。

XU Hong, born in 1967, Ph.D., professor. Her research interest is smart healthcare.



林卓胜(1991—),男,广东江门人,博士,副教授,研究方向为混沌密码、图像处理等。

LIN Zhuosheng, born in 1991, Ph.D., associate professor. His research interests include chaos cryptography, image processing, etc.



李胜可(1996—),女,河南汝州人,硕士研究生,主要研究方向为生物特征信息处理。

LI Shengke, born in 1996, M.S. candidate. Her research interest is biometric information processing.



刘慧琳(1997—),女,广东江门人,博士研究生,研究方向为计算机视觉、中医四诊客观化。

LIU Huilin, born in 1997, Ph.D. candidate. Her research interests include computer vision and objective study of four diagnosis in Chinese medicine.



冯跃(1959—),男,广东台山人,博士,教授,主要研究方向为计算机视觉、生物特征识别。

FENG Yue, born in 1959, Ph.D., professor. His research interests include computer vision and biometric features identification.