

基于非酒精性脂肪性肝病超声图像的预训练神经网络分类算法

郭丽娟 王文娟 王晓童 史莉玲

摘 要 超声检查因其无创性已成为诊断非酒精性脂肪性肝病(NAFLD)的首选方法,计算机辅助诊断技术的引入可以帮助医师减少在 NAFLD 检测和分类的偏差。为此,本研究提出了一种将结合注意力机制的预训练 VGG16 网络与 Stacking 集成学习模型相结合的混合模型,集合了基于自注意力机制的多尺度特征聚合和基于 Stacking 集成学习模型多分类模型(逻辑回归、随机森林、支持向量机)融合的特性,实现基于肝脏超声图像的正常肝脏、轻度脂肪肝、中度脂肪肝、重度脂肪肝的 4 分类,准确率为 91.34%,略优于传统神经网络算法($\leq 89.41\%$)。结果显示,相较于预训练 VGG16 网络,引入自注意力机制使得准确率提高了 3.02%,使用 Stacking 集成学习模型作为分类器进一步将准确率提高到 91.34%,超过了逻辑森林回归(89.86%)、支持向量机(90.34%)、随机森林(90.73%)等单一分类器。该方法能有效提升 NAFLD 超声图像检测的效率和准确性。

关键词 超声图像;非酒精性脂肪性肝病;VGG;自注意力;集成学习

[中图法分类号]R445.1;R575.5

[文献标识码]A

Pre-trained neural network classification algorithm based on ultrasound images of non-alcoholic fatty liver disease

GUO Lijuan, WANG Wenjuan, WANG Xiaotong, SHI Liling

Department of Ultrasound, Shanxi Children's Hospital, Shanxi Women and Children Hospital, Taiyuan 030013, China

ABSTRACT Ultrasound examination has become the preferred choice for diagnosing non-alcoholic fatty liver disease (NAFLD) due to its non-invasive. Computer-aided diagnosis technology can help doctors avoiding deviations of detection and classification in NAFLD. Therefore, this study propose a hybrid model that makes the pre-trained VGG16 network combined with the attention mechanism and the Stacking ensemble learning model, which has ability of multi-scale feature aggregation based on the self-attention mechanism and multi-classification model fusion (Logistic regression, random forest, support vector machine) based on Stacking ensemble learning. The proposed hybrid method achieves four classifications of normal, mild, moderate, and severe fatty liver based on ultrasound images, and it reaches an accuracy of 91.34%, which is slightly better than traditional neural network algorithms ($\leq 89.41\%$). The results show that compared with the pre-trained VGG16 network, adding the self-attention mechanism improves the accuracy by 3.02%. Using the Stacking ensemble learning model as a classifier further increases the accuracy to 91.34%, exceeding any one single classifier such as Logistic regression (89.86%), support vector machine (90.34%) and random forest (90.73%). The proposed hybrid method can effectively improve the efficiency and accuracy of NAFLD ultrasound image detection.

KEY WORDS Ultrasound image; Non-alcoholic fatty liver disease; VGG; Self-attention; Ensemble learning

近年来随着人们生活习惯的变化,脂肪肝的发病率呈逐年上升趋势,全球范围内患有非酒精性脂肪性肝病(non-alcoholic

fatty liver disease, NAFLD)的人群比例为 20%~30%,且趋于年轻化,给社会和家庭带来了沉重的医疗和经济负担。目前

基金项目:山西省基础研究计划(自由探索类)项目(20210302124433);山西省高等学校科技创新计划项目(2023L112)

作者单位:030013 太原市,山西省儿童医院 山西省妇幼保健院超声科(郭丽娟、王晓童、史莉玲);山西国际旅行卫生保健中心(王文娟)

通讯作者:史莉玲, Email: etyyck@126.com

NAFLD 的诊断和识别主要依赖于血液实验室检查和活组织病理检查等侵入性操作和超声、CT 等非侵入性成像技术^[1-2]。超声检查因其无创性和操作简便已成为诊断 NAFLD 的首选方法。然而,超声检查的准确性很大程度上取决于操作人员的技能和专业知识。近年计算机辅助诊断技术被引入以帮助临床医师准确诊断 NAFLD,并可以减少医师在 NAFLD 检测和分类应用中的偏差,有利于医院的资源优化和调度^[3-4]。在医学影像学技术中,基于机器学习的检测是目前研究热点之一。Andrade 等^[4]手动从超声图像中提取多个特征,并评估人工神经网络(ANN)、支持向量机(SVM)和 K 最邻近法(KNN)在诊断肝脏脂肪变性中的性能。Lupsor 等^[5]将图像的衰减系数应用到数理统计框架中以实现脂肪肝的检测。Das 等^[1]从不同受试者的超声图像中选择代表性感兴趣区(ROI)并提取纹理特征,然后评估如 SVM、神经网络和 XGBoost 等多种分类算法。Wu 等^[6]使用随机森林(RF)、朴素贝叶斯(NB)、ANN 和逻辑回归(LR)构建模型来预测 NAFLD。在上述机器学习算法中,特征选择在算法中起着至关重要的作用,对分类性能有着显著影响。然而,特征选择是由人工设计完成,其完整性和有效性依赖于研究人员的经验和知识。不同的特征选择方法可能影响最终的结果分析,而且当选择特征有误时可能会对结果产生负面影响。因此,制定更有效的策略来解决特征选择十分必要。

卷积神经网络(convolution neural network, CNN)可以实现特征的自动计算和选择,已经应用于超声图像分析中^[7-8]。Reddy 等^[9]应用 CNN 从超声图像中获取纹理特征,提高了 NAFLD 分类的诊断准确性。Che 等^[10]通过融合从二维超声图像数据提取的特征提出了一种新的多特征 CNN 架构,并提高了特征提取的性能。在基于 CNN 的架构中,针对特定的任务进行训练特定的 CNN 模型。在模型训练过程中,模型性能会随着层数的深度增加而提升,但深度模型的性能依赖于大量的样本数据、丰富计算资源,也容易出现过度拟合等问题。为了在使用深度神经网络优势的同时降低上述风险,可以采用在如 ImageNet 大数据集上的预训练 CNN,预训练的 CNN 仅充当特征提取器,无需依赖大量数据样本和计算资源进行重新训练,也可以避免过度拟合。Byra 等^[11]使用 Inception-ResNet-v2 架构进行肝脏超声图像特征提取。Zamanian 等^[12]将组合深度学习算法,即 Inception-ResNetV2、GoogleNet、AlexNet 和 ResNet101 应用于超声图像中的 NAFLD 分类,并结合了不同算法在性能上的优势。VGG 神经网络广泛应用于医学影像分析,在网络结构中采用 3×3 的卷积核并结合非线性层,多层结构增加网络深度,使网络具备较强的表征能力。Qu 等^[13]提出了一种基于 VGG 和 ViT 预训练结合模型,借助 VGG 实现了局部特征提取。呼延若曦等^[14]提出了一种以 VGG16 网络为编码器方式,实现了 MRI 图像中前列腺分区智能分割研究。在基于预训练的 CNN 的特征提取过程中,通常采用来自其深层特征,包含高级语义,缺少了诸如浅层的不同尺度信息。注意力机制作为神经网络应用中的关键技术之一,使其具备在不同网络层有选择性接受和处理信息的能力,提升了神经网络表征能力和分析、理

解能力^[15]。Li 等^[16]在设计的网络中引入了基于注意力机制的模块,在数据集上的准确性优于其他算法,性能有明显提升。宫霞等^[17]提出了将注意力机制引入基于级联结构的 UNet 网络,实现了对病灶边缘区域更加清晰的分隔。

此外,在特征提取之后,通常连接一个单一分类器实现对 NAFLD 的检测与分类。但单一的分类模型难以适应复杂的应用场景。集成学习算法将多个模型结果集成并作为最终结果,可以得到鲁棒性更强的结果,已被广泛应用于面向健康大数据诊疗领域中^[18]。An 等^[19]提出了一个深度集成学习框架应用于阿尔茨海默病的分类,提升了基分类器的多样性及分类准确性。熊思伟和刘玉琳^[20]提出基于 RF 模型、XGBoost 模型和 AdaBoost 模型等分类器的 Stacking 集成学习模型,实现对前列腺肿瘤风险预测性能的提升。因此,将预训练的 VGG、自注意力机制、Stacking 集成学习模型结合,能够融合预训练模型的性能、自注意力机制的强表征能力和 Stacking 集成学习模型的强鲁棒性,提升神经网络在 NAFLD 超声图像分类应用中的性能。本研究提出了一种将结合注意力机制的预训练的 VGG16 网络与 Stacking 集成学习模型相结合的混合模型,集合了基于自注意力机制的多尺度特征聚合和基于 Stacking 集成学习模型的多分类模型(LR、RF、SVM)融合的特性,实现基于肝脏超声图像的正常肝脏、轻度脂肪肝、中度脂肪肝、重度脂肪肝的 4 分类,同时使用准确度(ACC)、精确度(PRE)、召回率(REC)、F1 分数(F1-score)、曲线下面积(AUC)对模型性能进行评估。

一、方法和数据

本研究提出将结合注意力机制的预训练 VGG16 网络与 Stacking 集成学习模型相结合的混合模型,对图像处理过程主要包括图像预处理、特征提取和分类 3 个阶段:第一阶段,实施图像预处理,具体方法包括图像裁剪、降噪和数据增强,以改善图像质量,从而更清晰地观察图像特征;第二阶段,使用具有自注意力的预训练 VGG16 网络从超声图像中提取特征,具体方法为将原始超声图像输入到预训练 VGG16 网络,并从不同的卷积层输出不同的特征图,基于注意力的特征融合层连接并组合这些包含不同尺度信息的特征图;第三阶段,通过结合 LR、SVM 和 RF 等传统机器学习分类器,使用集成学习算法构建 Stacking 集成学习模型,使用基于自注意力的预训练 CNN 生成的特征,而非使用原始的超声图像作为原始输入,实现对 NAFLD 的预测。见图 1。

1. 超声图像采集

本研究数据来源于 2018 年 1 月至 2023 年 6 月山西省儿童医院经脂肪肝超声诊断标准确诊的 1964 个病例,每个病例选取肝右叶最大斜径切面,共获取 1964 张超声图像。由 2 名或以上主任医师对数据库中的超声图像进行评估,所有评估结果均被标记为 4 种类型之一,即正常肝脏、轻度脂肪肝、中度脂肪肝和重度脂肪肝;其中正常肝脏 571 例,轻度脂肪肝 517 例,中度脂肪肝 482 例,重度脂肪肝 394 例。当所有专家对同一病例的评估彼此一致时,评估被判断为有效且可接受。

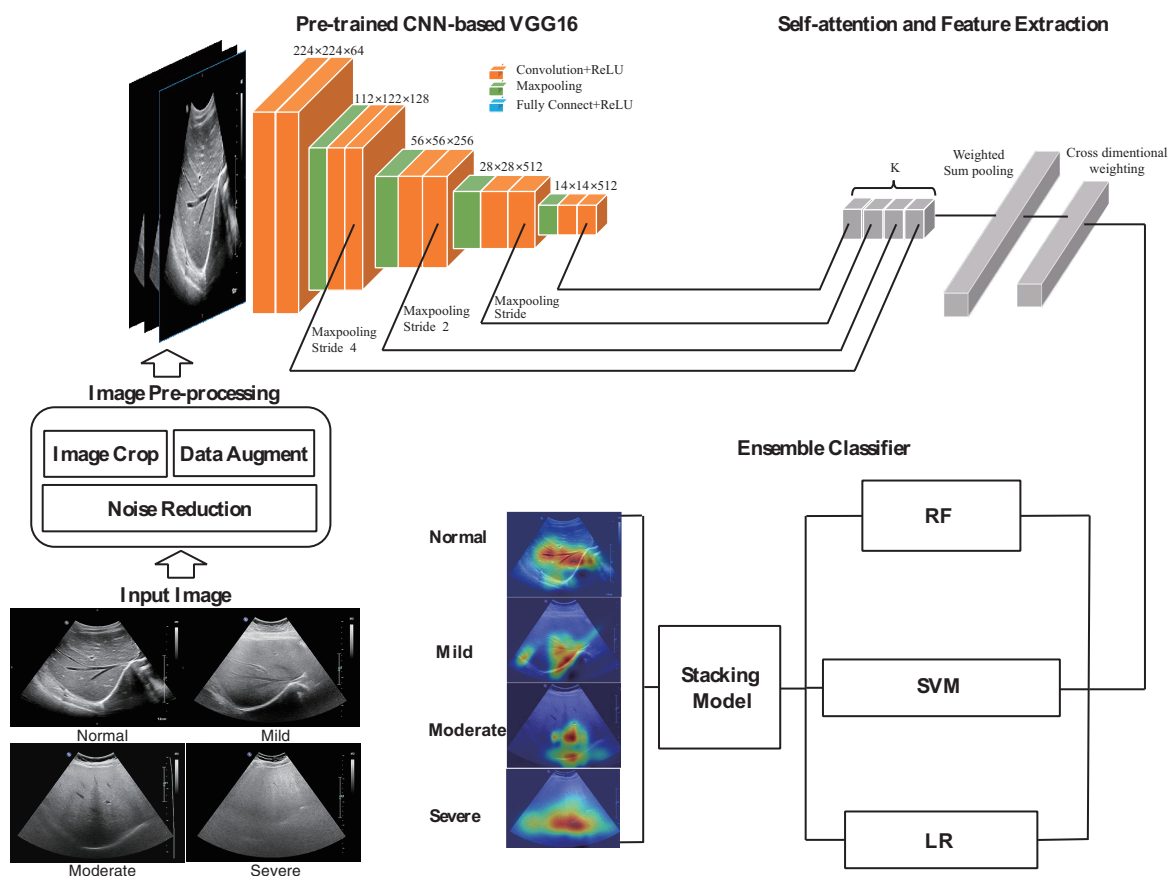


图1 基于注意力机制的预训练VGG16网络框架图

2. 图像预处理

由于设备等原因,超声图像在图像质量和尺寸均存在偏差,将会影响CNN架构的准确性和鲁棒性。为此,在算法的第一阶段进行了数据预处理,如图像裁剪、直方图均衡化、数据增强等。在超声图像中肝脏影像的两侧存在过多的空白或黑色空间且不传达任何诊断信息,因此所有图像均被统一裁剪。考虑到在预训练的VGG16网络中使用卷积层的输出,且模型中没有应用全连接层,故将裁剪后的图像调整为固定尺寸500×500,保证见效图像同一尺寸的同时也可以保留足够空间信息。同时引入直方图均衡化,通过对图像强度值进行变换以增强不同脂肪肝程度的超声图像对比度,有利于病灶的识别和分类。使用数据增强方式实现数据集扩展,如相对于y轴镜像图像、通过添加高斯噪声来调整图像中的照明,以及在(-30°, +30°)范围内旋转图像随机角度,可以避免神经网络过度拟合。数据集扩展到4000张超声图像,正常肝脏、轻度脂肪肝、中度脂肪肝、重度脂肪肝每个类别有1000张超声图像。在

算法模型学习和测试过程中,针对每一个类别随机选取20%患者的超声图像作为算法模型的测试集,并使用数据增强方式将数据扩充至200张超声图像;同样针对每一个类别随机选取80%患者的超声图像用于算法模型学习,并使用数据增强方式将数据扩充至800张超声图像。使用k折交叉验证进行训练验证,k为5,即随机选取学习数据集中的80%(640张超声图像)作为训练集,选取20%(160张超声图像)作为验证集。各数据集数据分布见表1。

3. 预训练的VGG16网络

CNN在特征提取方面具有良好的性能,将预训练的CNN作为新数据集的特征提取器是一个十分有效的措施。获得借助大数据集进行训练和层数更深的神经网络模型,在小规模的数据集也会有良好的表现。本研究选择的基于CNN的VGG16网络是在ImageNet数据集上使用超过120万张图像和1000个类别完成训练^[21-22]。基于CNN的VGG16网络结构是由13个具有非常小的感受野3×3的卷积层、5个大小为2×2的用于执行空间池

表1 各数据集数据分布

数据集	正常肝脏		轻度脂肪肝		中度脂肪肝		重度脂肪肝	
	病例	图像数(张)	病例	图像数(张)	病例	图像数(张)	病例	图像数(张)
训练集	365	640	331	640	308	640	252	640
验证集	91	160	83	160	77	160	63	160
测试集	115	200	103	200	97	200	79	200
合计	571	1000	517	1000	482	1000	394	1000

化的最大池化层(max pooling)、3个全连接层及1个softmax层组成。ReLU激活应用于所有隐藏层。VGG16网络结构大致可以分为5个卷积块(C1~C5),后面连接了3个全连接层(FC6~8)和1个softmax层,如图2所示。在每一层输出卷积特征图,包含不同尺度的图像信息。在接下来的步骤中,来自C2~C5阶段的特征图被聚合并输入到自注意力层,从而从图像中获得多尺度信息。

4. 自注意力和特征提取层

将预训练的VGG16网络中C2~C5阶段生成的特征图通过自注意力层聚合,生成具有更强表征能力的K维特征图。基于CNN的VGG16网络结构,将C2~C5阶段最后一层作为输出的特征向量图,用于后续自注意力机制特征融合。如图3所示,

C2~C5各个阶段输出特征向量图存在尺寸上的差异,通过形成相同尺寸大小的融合特征向量图,实现来自C2~C5不同阶段特征向量图的融合。其中,融合向量图的尺寸选择C4阶段的特征图大小作为目标尺寸,介于C2~C5阶段,使得层能够更平滑地传输特征,从而减少特征的损失,也有利于融合特征向量图更好地适应不同尺度下的特征^[23]。C2、C3阶段的特征图通过使用不同的步长从而完成特征图下采样过程,C5阶段的特征图采用双线性插值实现上采样,输出与C4阶段的特征图大小相同的融合特征向量图。

在多尺度特征图连接后,基于自注意力机制实现特征图聚合,其过程包括空间加权和通道加权,如图4所示。

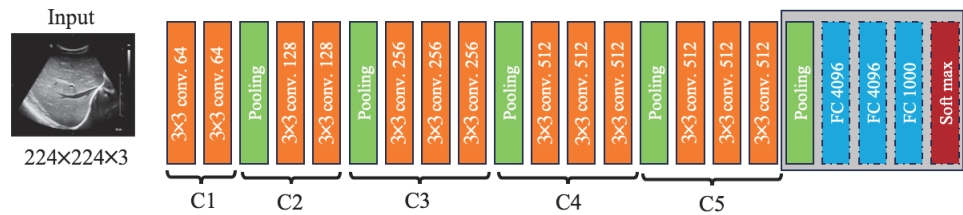


图2 VGG16网络的结构图

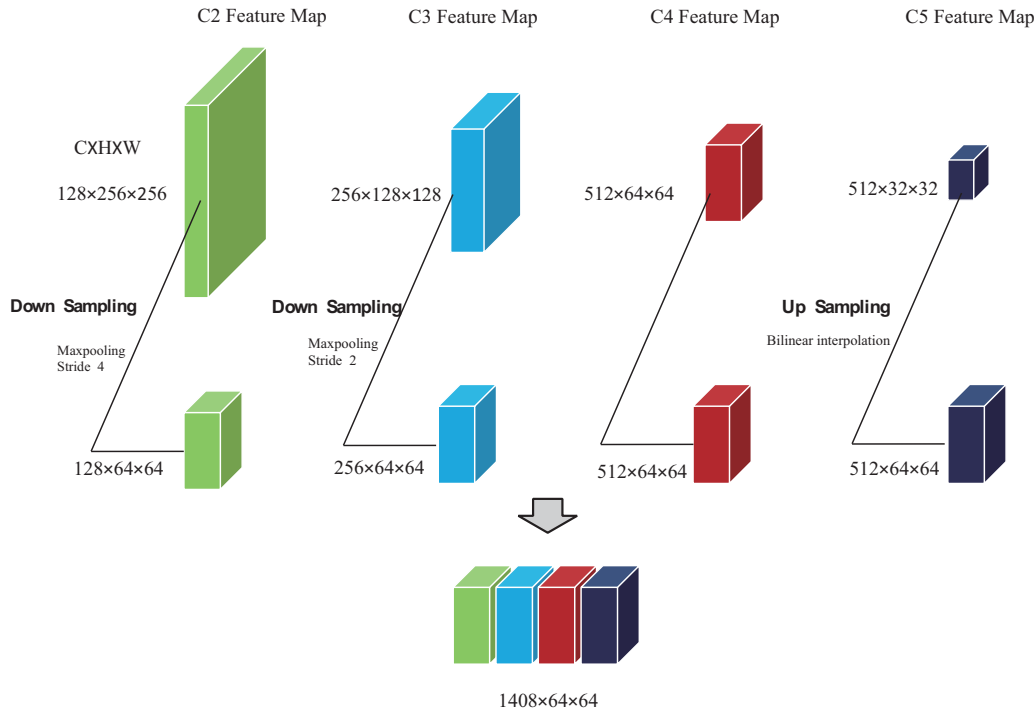


图3 多尺度特征图连接

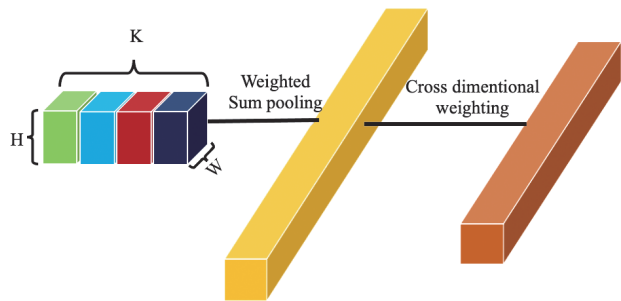


图4 基于自注意力的特征层聚合

(1)空间加权:与传统的求和池化^[24]不同,加权求和池化提供了一种基于特定位置重要性进行空间加权区分的机制,增强了复杂场景的表征能力。通过学习关注显著的病灶线索,其可以更好地发现NAFLD超声图像中的特征信息。在空间加权中,通过计算级联特征图中每个关键特征的相对比例以获得空间权重图S。将连接特征图上的像素比率作为无监督的空间权重,用于激活图形局部的丰富信息。在模型训练期间,其可以学习在无监督的情况下为特征更为明显的局部分配更高的权重。对于给定连接的特征图 $X \in \mathbb{R}^{H \times W \times K}$,其中H、W和K分别表示

通道的高度、宽度和通道数。多尺度特征图 $M \in \mathbb{R}^{h \times w}$ 表示为级联特征图 X 的通道维度上的总和,如公式(1)所示:

$$M = \sum_k X^k \quad (1)$$

其中, X^k 表示第 k 个通道的特征图。

空间权重图 S 由元素 S_{xy} 组成,即对应元素 M_{xy} 在所有元素之和中的比例; S_{xy} , 即空间坐标 (x, y) 的权重,其定义如公式(2)所示:

$$s_{xy} = \alpha \left(\frac{M_{xy}}{\left(\sum_{m,n} M_{m,n}^{1/a} \right)^a} \right)^b \quad (2)$$

其中, (m, n) 表示特征图上的位置。 a, b 为标准化参数,分别设定为 2, 0.5; α 是可学习的自适应系数,可由核大小为 1×1 的卷积获得。

最后,定义为加权求和池化的输出 Φ ,如公式(3)所示:

$$\Phi = \sum_{y=1}^H \sum_{x=1}^W S_{xy} f(x, y) \quad (3)$$

其中 $f(x, y)$ 是级联特征图 X 中相应位置 (x, y) 的值。空间加权增强了显著对象的重要性并减少了不显著对象的影响^[25-26]。

(2)通道加权:一般而言,激活特定类别的特定通道通常保持相对一致^[27]。这种一致的通道稀疏模式本身提供了有关类别判断的区分信息。为此,基于激活结构的稀疏性和权重大小实现了跨通道加权。通过计算每个通道的非零元素的属性以表征稀疏性。 Ω_k 代表每个通道的稀疏度,如公式(4)所示:

$$\Omega_k = \frac{1}{WH} \sum_{ij} \begin{cases} 1 & f(x, y)^{(k)} \geq 0 \\ 0 & f(x, y)^{(k)} < 0 \end{cases} \quad (4)$$

加权求和池化已经激活了包含 NAFLD 超声图像中常见特征的通道。然而,区分能力通常在于捕捉不常见的特定类别线索的通道。因此,本研究需增加表征出稀疏激活的通道的相对贡献,这些通道往往会编码不常见但信息丰富的特征。为了实现这一点,通道注意力机制的权重以高系数性为主,降低激活权重大小的贡献度。具体来说,通道 k 的权重 ω_k 定义为:

$$\omega_k = \beta \log \left(\frac{K + \sum_h \Omega_h}{\delta + \Omega_k} \right) \quad (5)$$

其中, ω_k 是通道 k 的权重, K 是级联特征图 X 的通道数, δ 是一个无限趋近于零的小数; β 是训练过程中可学习的自适应系数,可由核大小为 1×1 的卷积获得。

计算权重后,将空间加权图应用于每个通道的位置并计算每个通道的总和。因此,得到一个与通道数 K 相同的特征向量 Y ,如公式(6)所示:

$$Y = \omega_k \Phi = \beta \log \left(\frac{K + \sum_h \Omega_h}{\delta + \Omega_k} \right) \cdot \sum_{y=1}^H \sum_{x=1}^W S_{xy} f(x, y) \quad (6)$$

计算空间和通道权重后,将空间权重应用于每个通道的特

征图。然后对每个特征图中的空间位置加权求和,生成长度为 K 的通道特征向量。这将判别性空间信息聚合成为具有跨维度权重 ω_k 的紧凑跨通道表征。实际上,通道权重增加了信息丰富但不常见的视觉特征的相对贡献。通过调整模型以关注不常见的判别线索而非密集的无区别响应,跨通道加权增强了类型见差异。因此,基于激活稀疏模式的统计数据,强调特定类别的通道和空间位置的加权特征向量,使得模型能够利用多维信号来识别不同类别的 NAFLD。

5. Stacking 集成学习模型

Stacking 集成学习模型是通过组合多个分类器从而创建出一个新的分类模型,其考虑了异构的易受攻击、易错的学习者,在训练过程中,可以并行地学习该过程。Stacking 集成学习模型通过训练元分类器以组合优化一组易受攻击的基分类器,该元分类器根据易受攻击的基分类器的结果进行预测。相比任何其他基于个体的学习器,Stacking 集成学习模型提高了整体准确性^[26],其基本原理可以表示如公式(7)所示:

$$y^{SE} = \sum_{m=1}^M \omega_m y^m, \text{ 且 } \sum_{m=1}^M \omega_m = 1, \text{ 且 } \omega_m \geq 0 \quad (7)$$

其中, y^{SE} 表示 Stacking 集成学习模型的最终结果, y^m 表示基分类器, ω_m 表示元分类器中对应基分类器 y^m 的权重。

Stacking 集成学习模型的基本步骤为:①使用各个基分类器对故障样本进行识别和判断,获得多个结果;②将得到的多个结果作为训练样本输入到元分类器中训练和使用,获得最终的诊断结果。具体如图 5 所示。本研究中 Stacking 集成学习模型第一阶段采用 SVM、LR 和 RF 作为基分类器,每一个基分类器会产生各自的不同分类结果;在第二阶段采用 XGBoost 作为元分类器,使用第一个阶段中每个基分类器的分类结果作为输入,经过算法得出最终的分类结果。其中, SVM 是一种用于回归和分类分析的监督学习方法。其通过在 n 维空间中寻找跨类间隔最大的最佳超平面对数据进行分类,是一种用于回归和分类分析的监督学习方法^[27-28]。 LR 是一种基于概率原理实现的有监督机器学习算法,用于对分类或数值数据进行分类, LR 使用 sigmoid 函数将预测值映射到概率,其中 L2 正则化也称为岭正则化,用于 LR 模型的训练,以减少过度拟合并避免记忆。 RF 算法是一种监督机器学习算法,用于处理分类和回归问题^[29],其是由许多棵树组成的,随着树的增加, RF 算法增强了模型的随机性。 XGBoost 作为一种梯度提升树算法,其基本思想是通过不断地添加树拟合上次预测的残差,以提升模型的性能。

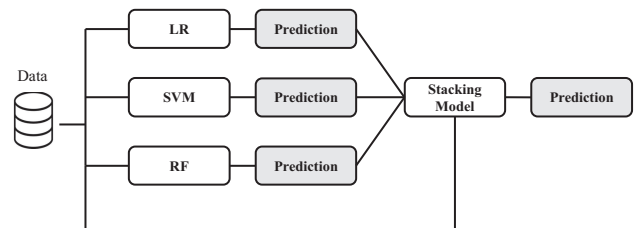


图 5 Stacking 集成学习模型框架

二、实验结果与分析

1. 评价指标

分类性能通过多个指标进行评估,包括 ACC、PRE、REC、F1-score 和 AUC。这些指标的计算取决于混淆矩阵,该矩阵用于计算真阳性(TP)、真阴性(TN)、假阳性(FP)和假阴性(FN),如表 2 所示。其中,TP 和 TN 分别表示正确预测的阳性样本和阴性样本的数量;FP 和 FN 分别表示错误预测的阳性样本和阴性样本的数量。混淆矩阵总结了分类模型的性能,每行代表预测的类别,每列表示真实的类别。矩阵中的值表示以其真实标签为条件,分配给每个预测类别的测试样本的百分比或原始计数。因此,对角线上的数据是预测类别与真实类别匹配的案例,显示了模型正确分类的情况;相反,非对角线值揭示了分类器的错误分类,即预测的标签与实际类别不一致。

表 2 混淆矩阵

预测值	真实值	
	阳性	阴性
阳性	TP	FP
阴性	FN	TN

TP:真阳性;TN:真阴性;FP:假阳性;FN:假阴性

ACC 表示观测值中,与所有样本相比,真实阳性和真实阴性的比例,其定义如公式(8)所示:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

PRE 是指标记为阳性的示例中实际为阳性的比例,其定义如公式(9)所示:

$$PRE = \frac{TP}{TP + FP} \quad (9)$$

REC,也称为灵敏度,表示实际阳性结果被正确预测为阳性的比例;与 PRE 相反,REC 衡量了模型识别数据中所有正面实例的能力。其计算方式如公式(10)所示:

$$REC = \frac{TP}{TP + FN} \quad (10)$$

F1-score 是平衡 PRE 与 REC 的机器学习模型的性能指标,计算方式为 PRE 和 REC 的调和平均值,对两种度量给予相同的权重,如公式(11)所示。F1-score 越高表示性能越好,因此在比较模型时,其通常用于原始 ACC 的替代方案。

$$F1-score = \frac{2 \times (PRE \times REC)}{PRE + REC} \quad (11)$$

2. 模型的超参数

学习率是一个超参数,决定神经网络在训练期间更新权重的步长,其影响模型学习最小化其损失函数的速度。在微调预训练模型时,冻结未重新训练的神经网络层,不需对其参数进行过多的更新。因此,应使用较正常训练更小的学习率进行微调,选择较小的学习率并将其设置为 0.0001。

mini-batch 大小决定了模型参数更新之前需处理样本的数量,较大的 mini-batch 需要更大的内存,并导致相同权重的训练周期更长,这可能会影响模型性能。因此,找到最佳 mini-batch 大小对于

最大化模型精度非常重要。本研究中 mini-batch 大小设置为 64。

深度学习中优化器的目标是更新偏差和权重参数以最小化成本或损失函数。针对给定问题选择最佳优化器可以通过适当调整权重和偏差值以提高模型性能并加速训练。本研究选择 Adam 作为优化器,用于提高所提出算法的准确性。

在训练期间,每个 epoch 均涉及使用数据集中的所有图像以更新网络权重以进行学习迭代。最佳的 epoch 数量取决于数据集大小、模型深度、学习率和优化器等因素。本研究所有数据集模型训练中 epoch 设为 100。

使用 Stacking 集成学习模型对训练集上基于注意力的 VGG16 网络的概率输出进行训练和优化,其中,Stacking 集成学习模型中的基分类器为 LR、SVM 和 RF,元分类器为 XGBoost。利用网格搜索来优化 Stacking 集成学习模型的超参数值,如下所示:

SVM: C=[0.1, 1, 100, 1000], gamma=[1, 0.1, 0.01, 0.001, 0.0001], kernel=RBF

RF: bootstrap=[True, False], max_depth=[10, 20, 30, 40, 50, 60], max_features=['auto', 'sqrt'], min_samples_leaf=[2, 5, 10], min_samples_split=[1, 2, 4]

LR: C=np. logspace(-3, 3, 7), penalty=['l2'], solver=['liblinear', 'newton-cg']

XGBoost: 'max_depth': 7, 'learning_rate': 0.15, 'min_child_weight': 2, 'n_estimators': 120, 'subsample': 0.8

通过网格搜索调整每个元学习器分类器的超参数值,用以优化所提出的算法在训练数据上的性能。

3. 实验结果

所提出的方法在 PyTorch 1.9.0 中实现,并使用主频为 2.30 GHz 的 Intel Core i7-10875H CPU 及包含 8 GB RAM 的 NVIDIA GeForce GTX 1070 GPU 进行评估。该硬件配置支持模型开发的快速迭代和数据集的高效批处理。

为了分析算法模型的性能,将 SVM、XGBoost、VGG19、ResNet101、InceptionV4 和 DenseNet161 等一些经典机器学习方法应用于对比实验,如表 3 所示。结果显示,尽管进行了广泛的超参数优化,但表 3 中的 CNN 在此数据集上的性能均偏中等。使用 DenseNet161 架构获得的 ACC 为 89.41%±0.21%,其余 CNN 模型的 ACC 集中在 87%~89%。本研究所提出的方法获得的 ACC 为 91.34%±0.16%,略高于其他算法。证实了本研究所提方法的有效性。

表 3 不同模型的衡量指标比较

模型	ACC(%)	F1-score	AUC
SVM	75.88±0.18	0.744	0.75
XGBoost	73.75±0.17	0.747	0.76
VGG19	87.04±0.24	0.869	0.88
ResNet101	87.29±0.16	0.877	0.90
InceptionV4	88.93±0.16	0.883	0.91
DenseNet161	89.41±0.21	0.887	0.90
本研究模型	91.34±0.16	0.907	0.92

ACC:准确度;F1-score:F1 分数;AUC:曲线下面积

为了验证提出的模型中每个部分的作用,本研究进行了一系列消融实验,结果显示,对比仅使用预训练的 VGG16 网络性能,增加了本研究提出的注意力机制的 VGG16 网络 ACC 提高了 3.02%,证明了本研究提出的自注意力机制在该任务中对提高模型性能发挥了重要作用。在连接了带有自注意力机制的预训练的 VGG16 网络后,分别使用 LR、SVM 和 RF 进行分类,结果显示 LR、SVM、RF 的 ACC 分别为 89.86%、90.34%、90.73%,RF 的 ACC 略高于 LR 和 SVM,但均低于 Stacking 集成学习模型在数据中实现的 ACC(91.34%)。见表 4。

表 4 消融实验结果

模型	ACC(%)	F1-score	PRE	REC
VGG16	86.28	0.871	0.872	0.868
VGG16+attention	89.36	0.891	0.890	0.893
VGG16+attention+LR	89.86	0.904	0.907	0.902
VGG16+attention+SVM	90.34	0.897	0.895	0.898
VGG16+attention+RF	90.73	0.902	0.901	0.903
VGG16+attention+Stacking 集成学习模型	91.34	0.907	0.908	0.905

ACC:准确度;F1-score:F1 分数;PRE:精确度;REC:召回率

在消融试验中,隔离了所提方法实现多尺度特征提取的注意力机制层和融合多分类器优势的 Stacking 集成学习模型,消除每一个关键部分时,ACC 的变化量化了其在统一框架内的重要性。

三、结论

本研究提出了一种将结合注意力机制的预训练的 VGG16 网络与 Stacking 集成学习模式相结合的混合模型,集合了基于自注意力机制的多尺度特征聚合和基于 Stacking 集成学习模型的多分类模型(LR、RF、SVM)融合的特性,实现了基于肝脏超声图像的正常肝脏、轻度脂肪肝、中度脂肪肝、重度脂肪肝的 4 分类,可应用于辅助 NAFLD 的超声图像检测,降低医师在 NAFLA 分类中的偏差。本研究的提出混合算法模型的 ACC 为 91.34%,略优于传统神经网络算法。结果表明,相较于预训练的 VGG16 网络,引入自注意力机制使得 ACC 提高了 3.02%,使用 Stacking 集成学习模型作为分类器进一步将 ACC 由 89.36% 提高到 91.34%。本研究所提出的算法模型为脂肪肝疾病的自动化监测和检测提供了可能性。

本研究提出的算法模型仍存在一些局限性:仅关注基于超声图像的脂肪肝分类,并未对病变进行定位。为了解决这些限制,今后工作将整合来自多个机构的更大数据库并引入区域分割算法以提高模型的实用性。随着模型和训练数据的进一步改进,该方法最终可能转化为检测和监测脂肪肝疾病的自动化工具。

参考文献

[1] Das A, Connell M, Khetarpal S. Digital image analysis of ultrasound images using machine learning to diagnose pediatric nonalcoholic fatty liver disease[J]. Clin Imaging, 2021, 77(1):62-68.

[2] Zhang L, Zhu H, Yang T. Deep Neural Networks for fatty liver ultrasound images classification [C]//2019 Chinese Control and Decision Conference(CCDC). IEEE, 2019:4641-4646.

[3] Cao W, An X, Cong L, et al. Application of deep learning in quantitative analysis of 2-dimensional ultrasound imaging of nonalcoholic fatty liver disease[J]. J Ultrasound Med, 2020, 39(1): 51-59.

[4] Andrade A, Silva JS, Santos J, et al. Classifier approaches for liver steatosis using ultrasound images[J]. Procedia Technol, 2012, 5(1): 763-770.

[5] Lupşor M, Badea R, Vicaş C, et al. Non-invasive steatosis assessment in NASH through the computerized processing of ultrasound images: attenuation versus textural parameters [C]//2010 IEEE International Conference on Automation, Quality and Testing, Robotics(AQTR). IEEE, 2010:1-6.

[6] Wu CC, Yeh WC, Hsu WD, et al. Prediction of fatty liver disease using machine learning algorithms [J]. Comput Methods Programs Biomed, 2019, 170(1):23-29.

[7] 沈筱梅, 张新颖, 王权泳, 等. 基于卷积神经网络的涎腺肿瘤超声图像分类研究[J]. 临床超声医学杂志, 2023, 25(10):849-855.

[8] 莫莹君, 刘友员, 郭瑞斌. 基于深度神经网络的颈动脉斑块超声图像诊断技术研究[J]. 临床超声医学杂志, 2022, 24(5): 382-385.

[9] Reddy DS, Bharath R, Rajalakshmi P. Classification of nonalcoholic fatty liver texture using convolution neural networks[C]//2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom). IEEE, 2018:1-5.

[10] Che H, Brown LG, Foran DJ, et al. Liver disease classification from ultrasound using multi-scale CNN[J]. Int J Comput Assist Radiol Surg, 2021, 16(9):1537-1548.

[11] Byra M, Styczynski G, Szmigielski C, et al. Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images[J]. Int J Comput Assist Radiol Surg, 2018, 13(12): 1895-1903.

[12] Zamanian H, Mostaar A, Azadeh P, et al. Implementation of combinational deep learning algorithm for non-alcoholic fatty liver classification in ultrasound images[J]. J Biomed Phys Eng, 2021, 11(1):73-84.

[13] Qu X, Lu H, Tang W, et al. A VGG attention vision transformer network for benign and malignant classification of breast ultrasound images[J]. Med Phys, 2022, 49(9):5787-5798.

[14] 呼延若曦, 吴哲, 许杉杉, 等. 基于 Vgg16-Unet 模型的 MRI 图像下前列腺分区智能分割研究[J]. 陆军军医大学学报, 2023, 45(13): 1441-1449.

[15] 张直政. 神经网络的注意力机制研究[D]. 合肥:中国科学技术大学, 2021.

[16] Li L, Xu M, Wang X, et al. Attention based glaucoma detection: a large-scale database and CNN model[C]. Proceedings of the IEEE/ CVF Conference on Computer Vision and Pattern Recognition. 2019:

- 10571-10580.
- [17] 宫霞,赵富强,吴卫华.级联结构网络在肺癌患者颈部淋巴结超声图像分割中的应用价值[J].临床超声医学杂志,2022,24(8):635-639.
- [18] 刘帅.健康大数据诊疗问题的集成学习算法研究与应用[D].长春:吉林大学,2023.
- [19] An N, Ding H, Yang J, et al. Deep ensemble learning for Alzheimer's disease classification[J]. J Biomed Inform, 2020, 105: 103411.
- [20] 熊思伟,刘玉琳.基于 Borderline-SMOTE 算法与 Stacking 集成学习的前列腺肿瘤风险预测研究[J].现代肿瘤医学,2023,31(16):3075-3081.
- [21] Fang Z, Wang Y, Peng L, et al. Integration of convolutional neural network and conventional machine learning classifiers for landslide susceptibility mapping[J]. Comput Geosci, 2020, 139: 104470.
- [22] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv: 1409.1556, 2014.
- [23] Zhang YM, Hsieh JW, Lee CC, et al. SFPN: Synthetic FPN for object detection[C]//2022 IEEE International Conference on Image Processing (ICIP). IEEE, 2022: 1316-1320.
- [24] Babenko A, Lempitsky V. Aggregating local deep features for image retrieval[C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 1269-1277.
- [25] Kalantidis Y, Mellina C, Osindero S. Cross-dimensional weighting for aggregated deep convolutional features[C]//Computer Vision—ECCV 2016 Workshops: Amsterdam, the Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 685-701.
- [26] Deng L, Yu D, Platt J. Scalable Stacking and learning for building deep architectures[C]//2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2012: 2133-2136.
- [27] Fawcett T. An introduction to ROC analysis[J]. Pattern Recogn Lett, 2006, 27(8): 861-874.
- [28] Chidambaram S, Srinivasagan KG. Performance evaluation of support vector machine classification approaches in data mining[J]. Cluster Comput, 2019, 22(1): 189-196.
- [29] Shi T, Horvath S. Unsupervised learning with random forest predictors[J]. J Comput Graphical Statistics, 2006, 15(1): 118-138.

(收稿日期:2023-12-04)

• 病例报道 •

Ultrasonic diagnosis of right cervical tracheal diverticulum: a case report 超声诊断右侧颈部气管憩室 1 例

魏凯能 张跃鹏

[中图分类号]R445.1

[文献标识码]B

患者男,52岁。4年前检出甲状腺结节,无颈部疼痛、声音嘶哑等不适;慢性阻塞性肺疾病(COPD)病史4年余。本次常规超声复查甲状腺结节:右下颈深方可见一大小2.3 cm×1.9 cm×1.3 cm气体样强回声包块,位于气管右侧及颈动脉间,与甲状腺分界清晰,上界达甲状腺右侧叶下极后方,下界因位置较深显示不清晰,该包块前界尚清晰,两侧及后界欠清晰,内侧边缘可见弧形低回声,厚薄不一,其内未探及血流信号,吞咽后及吞咽过程中其形态及超声表现均未见明显改变。见图1。超声提示:右下颈深方含气样包块,考虑气管憩室可能。颈胸部CT检查:气管右后侧局限性含气腔影,呈多分隔状,与气管相通,通道宽约1.2 cm,考虑气管憩室。见图2。因患者无明显临床症

状,未行外科干预,建议随诊。

讨论:气管憩室是一种气管旁含气囊肿,该病相对罕见,发病率为2.4%。患者通常无症状,多在胸部X线或CT检查中偶然发现^[1]。因右侧颈部背侧无食管支撑,而多见于气管右侧(97.1%),少见气管左侧(2.9%)^[2]。气管憩室可分为先天性和获得性,先天性气管憩室多见于男性,病变内径及与其气管连接口均较小,多位于声带下方4~5 cm处或隆突上方,常由气管发育异常引起^[3]。而获得性气管憩室可出现在任何水平,常为宽口,较先天性大,多见于气管壁薄弱合并慢性咳嗽或COPD患者。结合本例患者COPD病史、超声表现(包块体积较大)及胸部CT表现(憩室通道口较大)等,考虑获得性气管憩室可能

基金项目:武汉大学中南医院医学科技创新平台建设支撑面上项目(PTXM2022034)

作者单位:430071 武汉市,武汉大学中南医院综合超声医学科

通讯作者:张跃鹏,Email:z-yypeng@163.com