



(12) 发明专利申请

(10) 申请公布号 CN 116246654 A

(43) 申请公布日 2023. 06. 09

(21) 申请号 202310104612.4

(22) 申请日 2023.02.13

(71) 申请人 南昌大学

地址 330000 江西省南昌市红谷滩新区学府大道999号

(72) 发明人 张明辉 孙威威 孙萍 张帆

(74) 专利代理机构 北京众合诚成知识产权代理有限公司 11246

专利代理师 王焕巧

(51) Int. Cl.

G10L 25/18 (2013.01)

G10L 25/30 (2013.01)

G10L 25/66 (2013.01)

G10L 19/26 (2013.01)

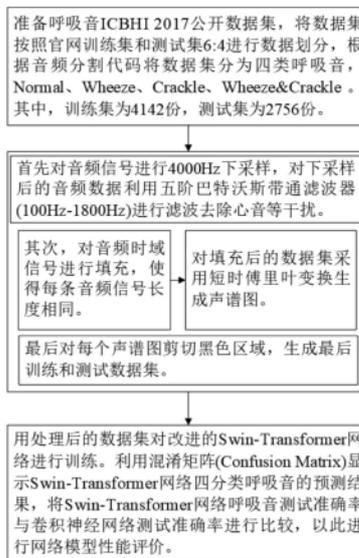
权利要求书1页 说明书6页 附图3页

(54) 发明名称

一种基于改进的Swin-Transformer的呼吸音自动分类方法

(57) 摘要

本发明公开了一种基于改进的Swin-Transformer的呼吸音自动分类方法,属于音频信号识别领域,包括三个步骤。步骤A:准备ICBHI 2017数据集,分为四类呼吸音:正常呼吸音、哮鸣音、爆裂音、哮鸣音和爆裂音。步骤B:音频信号预处理。首先对音频信号进行下采样并采用五阶巴特沃斯带通滤波器滤除心音等干扰。其次,对音频信号进行智能填充。最后采用短时傅里叶变换生成声谱图同时剪切黑色区域生成数据集。步骤C:采用训练数据对改进的Swin-Transformer网络进行训练。利用混淆矩阵显示Swin-Transformer网络四分类呼吸音的预测结果。本发明为呼吸音自动分类提供了一种新的改进方案,有利于促进呼吸音识别设备的发展。



1. 一种基于改进的Swin-Transformer的呼吸音自动分类方法,其特征在于,包括:

S1、准备呼吸音ICBHI 2017公开数据集,将数据集按照官网训练集和测试集6:4进行数据划分,根据音频分割代码将数据集分为四类呼吸音:正常呼吸音Normal、哮鸣音Wheeze、爆裂音Crackle、哮鸣音和爆裂音Wheeze&Crackle;

S2、音频信号预处理;首先对音频信号进行4000Hz下采样,对下采样后的音频数据利用五阶巴特沃斯带通滤波器进行滤波去除干扰;其次,对音频时域信号进行填充,使得每条音频信号长度相同并对填充后的数据集采用短时傅里叶变换生成声谱图;最后对每个声谱图剪切黑色区域,生成最后训练数据集和测试数据集;

S3、用处理后的数据集对基于改进的Swin-Transformer网络构建的分类模型进行训练;利用混淆矩阵显示Swin-Transformer网络四分类呼吸音的预测结果。

2. 根据权利要求1所述的基于改进的Swin-Transformer的呼吸音自动分类方法,其特征在于,在所述S2步骤中的呼吸音预处理具体过程包括:

S201、设置短时傅里叶变换FFT点数为256,其他参数默认,生成声谱图。

3. 根据权利要求1所述的基于改进的Swin-Transformer的呼吸音自动分类方法,其特征在于,在所述S3步骤中的Swin-Transformer测试具体过程包括:

S301、改进的Swin-Transformer网络测试部分,在呼吸音分类任务中,加入混淆矩阵作为网络的评价指标。

一种基于改进的Swin-Transformer的呼吸音自动分类方法

技术领域

[0001] 本发明属于音频信号识别领域,具体为一种基于改进的Swin-Transformer的呼吸音自动分类方法。

背景技术

[0002] 呼吸系统疾病是高发和高致死率的疾病,根据世界卫生组织发布的关于全球十大死亡疾病数据显示,每年与呼吸相关疾病的致死率占10%以上。早发现、早诊断、早治疗具有积极的临床意义。呼吸音听诊是呼吸系统疾病(特别是支气管炎、肺炎、支气管哮喘等)前期诊断最重要、最便捷的诊断手段。然而,使用传统双耳听诊器进行听诊受到很多限制:①听诊需要有经验的医生进行;②是一个主观的过程,听诊结果取决于听诊人感知、区分不同病理声音的经验和能力,容易造成漏诊、误诊;③难以进行定量测量以及检查记录难以保留和回放;④佩戴个人防护装备的医生很难在不违反防护标准的情况下进行听诊,限制了对患有通过空气传播或飞沫传播的肺部疾病患者的听诊。呼吸音自动分析有望克服这些限制。然而,异常呼吸音的自动识别和分类是一项具有挑战性的工作,因为呼吸音具有非平稳的性质,且与同一种疾病相关的异常呼吸音变异较大,不同患者的同种异常呼吸音特征可能不同,甚至单个患者在不同的呼吸周期内也可能发生改变。提高呼吸音自动分类准确率一直以来都是呼吸音分类领域亟待解决的问题。

发明内容

[0003] 针对上述提出的问题,本发明的目的是提出一种切实可行、效果优良的呼吸音自动分类方法,包括呼吸音数据预处理技术和基于改进的Swin-Transformer的呼吸音自动分类模型。

[0004] 为达到上述目的,本发明提供的技术方案具体如下:

[0005] 步骤A:准备呼吸音ICBHI 2017公开数据集,将数据集按照官网训练集和测试集6:4进行数据划分,根据音频分割代码将数据集分为四类呼吸音:正常呼吸音(Normal)、哮鸣音(Wheeze)、爆裂音(Crackle)、哮鸣音和爆裂音(Wheeze&Crackle)。其中,训练集为4142份,测试集为2756份;

[0006] 步骤B:音频信号预处理。首先对音频信号进行4000Hz下采样,对下采样后的音频数据利用五阶巴特沃斯带通滤波器(100Hz-1800Hz)进行滤波去除心音等干扰。其次,对音频时域信号进行填充,使得每条音频信号长度相同并对填充后的数据集采用短时傅里叶变换生成声谱图。最后对每个声谱图剪切黑色区域,生成最后训练数据和测试数据;

[0007] 步骤C:用处理后的数据集对基于改进的Swin-Transformer网络构建的分类模型进行训练。利用混淆矩阵(Confusion Matrix)显示Swin-Transformer网络四分类呼吸音的预测结果,将Swin-Transformer网络呼吸音测试准确率与卷积神经网络呼吸音测试准确率进行比较,以此进行网络模型性能评价。

[0008] 本发明提高了呼吸音分类准确率。具体而言,本发明通过对音频信号下采样,采用

巴特沃斯带通滤波器滤除干扰信号,利用智能填充技术和黑色区域剪切技术,使得每一类音频信号的声谱图具有比较明显的特征,对基于改进的Swin-Transformer分类网络模型进行训练。实验结果表明,本设计方案呼吸音分类效果显著。

[0009] 本发明具有以下技术优点:

[0010] (1)对原始音频数据按照特征出现时间提取相对应的特征片段,更加具有准确性。

[0011] (2)通过对音频数据采用智能填充技术,黑色区域剪切技术和短时傅里叶变换特征提取方案,更加具有可靠性。

[0012] (3)将基于改进的Swin-Transformer网络应用于呼吸音自动分类任务,并将混淆矩阵应用于网络测试中,更加具有先进性。

[0013] (4)将卷积神经网络分类模型和基于改进的Swin-Transformer网络模型进行比较,突显出基于改进的Swin-Transformer网络分类方法在呼吸音分类任务中优于卷积神经网络分类方法。

[0014] 下面结合附图和具体实施例,对本发明作更为详细的说明。

附图说明

[0015] 图1是基于改进的Swin-Transformer的呼吸音自动分类方法的流程图;

[0016] 图2是基于改进的Swin-Transformer的网络训练流程图;

[0017] 图3是基于改进的Swin-Transformer的呼吸音自动分类方法的总体结构框架图;

[0018] 图4是基于改进的Swin-Transformer混淆矩阵在2756个测试数据的结果图。

具体实施方式

[0019] 为详细说明本发明技术方案的技术内容,构造特征,所实现目的及效果。以下结合附图,对本发明进行进一步的详细说明。此处所描述的具体实施仅用于解释本发明技术方案,并不限于本发明。

[0020] 图1是本发明的流程图,结合附图1对本发明方法的具体步骤描述如下:

[0021] 步骤A:准备呼吸音ICBHI 2017公开数据集,将数据集按照官网训练集和测试集6:4进行数据划分,根据音频分割代码将数据集分为四类呼吸音:Normal、Wheeze、Crackle、Wheeze&Crackle。其中,训练集为4142份,测试集为2756份;

[0022] 步骤B:音频信号预处理。首先对音频信号进行4000Hz下采样,对下采样后的音频数据利用五阶巴特沃斯带通滤波器(100Hz-1800Hz)进行滤波去除心音等干扰。其次,对音频时域信号进行填充,使得每条音频信号长度相同并对填充后的数据集采用短时傅里叶变换生成声谱图。最后对每个声谱图剪切黑色区域,生成最后训练数据和测试数据;

[0023] 步骤C:用处理后的数据集对基于改进的Swin-Transformer网络构建的分类模型进行训练。利用混淆矩阵(Confusion Matrix)显示Swin-Transformer网络四分类呼吸音的预测结果,将Swin-Transformer网络呼吸音测试准确率与卷积神经网络呼吸音测试准确率进行比较,以此进行网络模型性能评价。

[0024] 本发明提高了呼吸音分类准确率。具体而言,本发明通过对音频信号下采样,采用巴特沃斯带通滤波器滤除干扰信号,利用智能填充技术和黑色区域剪切技术,使得每一类音频信号的声谱图具有比较明显的特征,对基于改进的Swin-Transformer分类网络模型进

行训练。实验结果表明,本设计方案呼吸音分类效果显著。

[0025] 上述步骤A中音频信号处理的具体过程包括:

[0026] A101、对ICBHI2017呼吸音公开数据按四类呼吸音进行音频时间片分割;

[0027] A102、将数据集按照官方给出的6:4比例分为训练集和测试集,并分类保存。

[0028] 在上述实施方式步骤A中,该数据库包含总共5.5小时的录音。包含6898次呼吸周期,其中3642个正常呼吸音样本,886个哮鸣音样本,1864个爆裂音样本,506个同时包含哮鸣音和爆裂音样本,来自126名受试者,920个注释音频样本。该呼吸音样本被分为:Normal、Wheeze、Crackle、Wheeze&Crackle四种类别。录音采用异构设备收集,每段时长为10s~90s不等的带注释的呼吸音音频。

[0029] 上述步骤B中,参照图2所示,呼吸音音频数据预处理部分具体步骤包括:

[0030] B101、对音频信号进行下采样频率4000Hz,采用五阶巴特沃斯带通滤波器100Hz~1800Hz对音频信号滤波;

[0031] B102、采用智能填充技术。复制同一周期音频时域信号,使得音频数据的长度保持一致;

[0032] B103、采用黑色区域剪切技术。经过巴特沃斯带通滤波生成的声谱图会存在黑色且无明显有效纹理的区域,采用该方法剪切黑色区域;

[0033] B104、设置短时傅里叶变换FFT点数为256,其他参数默认,生成声谱图。

[0034] 在步骤B101中,采用带通滤波器对音频信号进行滤波,由于巴特沃斯滤波器具有在通频带内的频率响应曲线最大平坦,没有起伏,而在阻频带则逐渐下降为零的优点,故本发明采用五阶巴特沃斯带通滤波器对音频信号进行滤波。

[0035] 在步骤B102中,采用智能填充技术。在ICBHI数据集中,呼吸周期的长度从0.2s到16.2s不等,平均周期长度为2.7s。然而,网络训练期望输入的数据长度保持一致。处理该问题的标准方法是通过零填充将音频信号填充到一个固定大小。受零填充思想启发,本发明采用新的智能填充方案。具体来说,对音频时域信号复制同一周期填充,继续这个过程,直到达到一致音频数据大小。一个小的音频样本数据在经过剪切以后,在一个已经很稀缺的数据集中失去了有价值的信息,而一个非常大的样本数据则产生了很多次的重复,降低网络分类性能。

[0036] 在步骤B103中,采用黑色区域剪切技术。在使用热力图分析卷积神经网络对呼吸音提取特征规律时,发现网络对声谱图有明显黑色区域的高频区域(1500Hz~1800Hz)不敏感。为避免对网络性能产生不利的影 响,有选择地从这些声谱图的高频率区域剪掉黑色区域。确保网络集中在感兴趣的区域,提高网络分类性能。

[0037] 在步骤B104中,采用短时傅里叶变换对音频数据进行处理,具体来说,采用加窗方法对每段时域信号进行傅里叶变换。在数字滤波器设计和频谱估计中,窗函数的选择对于整体结果的质量有重大影响。加窗主要作用是减弱因无穷级数截断而产生吉布斯现象的影响,根据窗函数和采样点数的大小,声谱图可分为窄带声谱图和宽带声谱图。宽带声谱图由于窗短,时间分辨率高、频率分辨率低,而窄带声谱图则由于窗长,时间分辨率低、频率分辨率高。

[0038] 根据声谱图易受窗宽影响的特点,对原始信号做短时傅里叶变换的步骤包括:

[0039] 首先将窗口移动到信号的开端位置,此时窗函数的中心位置在 $t = \tau_0$ 处,对信号加

窗处理

[0040] $y(t) = x(t) * w(t - \tau_0)$

[0041] 然后进行傅里叶变换

[0042]
$$X(\omega) = F(y(t)) = \int_{-\infty}^{+\infty} x(t) * w(t - \tau_0) e^{-j\omega t} dt$$

[0043] 得到第一个分段序列的频谱分布 $X(\omega)$ 。在实际应用中,由于信号是离散的点序列,因此经傅里叶变换得到频谱序列 $X[N]$,为了更加直观表示,定义函数 $S(\omega, \tau)$,表示在窗函数中心为 τ 时,对原函数进行变换后的频谱结果为 $X(\omega)$,公式表示为

[0044]
$$S(\omega, \tau) = F(x(t) * w(t - \tau_0)) = \int_{-\infty}^{+\infty} x(t) * w(t - \tau_0) e^{-j\omega t} dt$$

[0045] 对应离散情形下, $S[\omega, \tau]$ 是一个二维矩阵,每一列代表在不同位置对信号加窗,对得到的分段进行傅里叶变换后的结果序列。以此类推,每次移动窗函数一段距离,该距离一般小于窗口宽度,两窗之间存在一定的重叠部分,重复以上操作,最终得到 $\tau_0 - \tau_N$ 上所有分段的频谱结果,用 S 表示。

[0046] 上述步骤C中,参照图3所示,改进Swin-Transformer网络测试部分具体步骤包括:

[0047] C101、改进Swin-Transformer网络测试部分,未改进的Swin-Transformer网络测试部分为输出单一的分类准确度,此方法对网络性能的评价不够具体。因此,在呼吸音分类任务中,加入混淆矩阵作为网络的评价指标。不仅可以更好的看出每一类呼吸音的准确度,而且能够看出网络对呼吸音数据集训练效果。

[0048] 在上述实施方式中,步骤C101中,采用混淆矩阵评价网络训练效果,原始网络评价网络性能只能输出单一的训练测试准确度,无法看出每一类的预测效果。在医学领域的预测和诊断中,这样的结果显然是不够优越的。利用混淆矩阵对网络性能进行评价,不仅可以看出每一类的准确度,同时也能计算出模型的敏感性和特异性,有利于呼吸音分类任务的预测和诊断。混淆矩阵包含了实际分类和预测分类的信息,通过对包含的信息进行选择计算,可以估计网络训练的分类能力。下表展示了混淆矩阵具体结构。

[0049]

混淆矩阵	网络预测是正类	网络预测是负类
样本实际是正类	TP	FN
样本实际是负类	FP	TN

[0050] 上表显示了 $n=2$ 的混淆矩阵,与网络分类相关的大小为 $n \times n$ 的混淆矩阵显示了预测和实际分类结果。其中 n 是不同类别的数量,TP指样本真实类别是正类,网络预测结果是正类,FP指样本真实类别是负类,网络预测结果是正类,TN指样本真实类别是负类,网络预测结果是负类,FN指样本真实类别是正类,网络预测结果是负类。预测精度和分类误差可以从这个矩阵中得到。具体为,准确率 $Score = (TP+TN) / (TP+FP+FN+TN)$,敏感度 $Sensitivity(SE) = TP / (TP+FN)$,特异度 $Specificity(SP) = TN / (TN+FP)$,F1-值(F1-score) $= 2 * TP / (2 * TP + FP + FN)$ 。其中,敏感度指样本真实类别是正类的结果中,网络预测是正类的比重,特异度指样本真实类别是负类的结果中,网络预测是正类的比重。当FN或FP为0时,表明网络中类别预测分歧为1,在这种情况下,混淆矩阵只对一个类别进行错误的分类。当FN或FP相同时,表明网络中类别预测分歧为0。FN或FP不仅本身具有良好的辨别力,更重要的

是相互之间具有互补性。基于改进的Swin-Transformer混淆矩阵在2756个测试数据的结果图如图4所示,混淆矩阵应用到Swin-Transformer不仅提高了网络性能的辨识度,同时能够清晰展示每种呼吸音类别和Swin-Transformer网络对呼吸音数据集的整体预测效果,对于Swin-Transformer网络应用到其他分类任务具有非常好的参考价值。

[0051] 本发明提出基于Swin-Transformer的呼吸音分类方法,其实验实施配置要求如下:

[0052] 本方法已经在python中实现了所提出的架构,Intel Core i7-6800中央处理器和NVIDIA Geforce GTX 2080Ti显卡的台式计算机上运行了所有实验。对于该实验的配置要求如下,深度学习框架需为Pytorch,模型训练时显卡显存需大于等于11264MB,测试时无显卡性能要求。数据集为ICBHI 2017呼吸音数据集,原始数据集中包括3642个正常呼吸音样本,886个哮鸣音样本,1864个爆裂音样本,506个同时包含哮鸣音和爆裂音样本。应用改进Swin-Transformer对经过下采样、五阶巴特沃斯带通滤波、智能填充技术、黑色区域剪切技术和短时傅里叶变换处理生成的数据集进行分类实验。

[0053] 表1.不同网络对呼吸音四分类准确度比较

方法	准确率
Multi-images+VGG16	47%
LungRN+NL	52.3%
Prototype Learning	55.4%
CNN	55.2%
CNN+CBA+BRC	55.7%
CNN+CBA+BRC+FT	56.2%
Improvement Swin-Transformer	57.3%

[0054] 实验结果如表1所示,实验表明,在对ICBHI2017呼吸音数据集四分类预测实验中,基于改进Swin-Transformer网络取得了57.3%的分类准确率,高于最新传统卷积神经网络的测试性能。本发明对于提高呼吸音分类准确率具有较为明显的效果。

[0055] 表2.不同网络对呼吸音四分类混淆矩阵评价对比结果

	方法	SP	SE	Score
	Jakoveljevic <i>et al.</i>	--	--	39.5%
	Chambres <i>et al.</i>	78.1%	20.8%	49.4%
	Serbes <i>et al.</i>	--	--	49.9%
[0057]	Ma <i>et al.</i>	69.2%	31.1%	50.2%
	Ma <i>et al.</i>	63.2%	41.3%	52.3%
	CNN	71.4%	39.0%	55.2%
	CNN+CBA+BRC	71.8%	39.6%	55.7%
	CNN+CBA+BRC+FT	72.3%	40.1%	56.2%
[0058]	Improvement Swin-Transformer	74.1%	40.4%	57.3%

[0059] 表2为不同网络对呼吸音四分类评价对比结果。其中,SE的计算公式为 $SE = TP_i / [(TP_c + FN_c) + (TP_w + FN_w) + (TP_b + FN_b)]$, ($i \in \{\text{Crackle}; \text{Wheeze}; \text{Crackle\&Wheeze}\}$), SP的计算公式为 $SP = TP_n / (TP_n + FN_n)$, 下标简写字母分别为{Normal (n), Crackle (c), Wheeze (w), [Crackle&Wheeze] (b)}, Score即为预测正确样本占所有测试样本的比例。从表中可以看出,本发明分类准确率明显高于传统卷积神经网络分类结果。表明本发明具有良好的分类性能,由于该情况符合实际场景,因此本发明具有重大研究意义。

[0060] 以上所述仅表达了本发明的优选实施方式,其描述较为详细和具体,但不能因此而理解为对本发明专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本发明构思的前提下,还可以做出若干变形、改进以及替代,这些都属于本发明的保护范围。因此,本发明的保护范围应以所附权利要求为准。



图1



图2

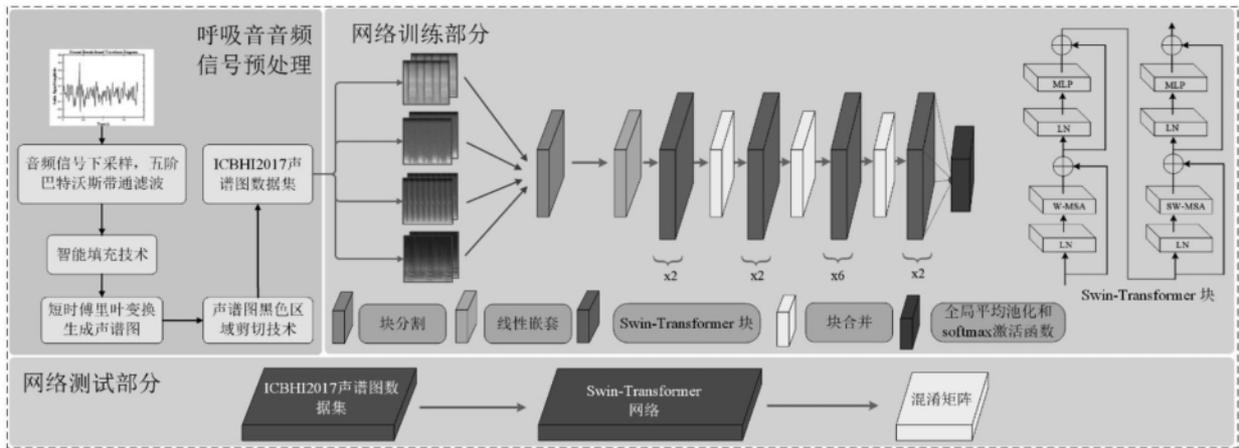


图3

混淆矩阵

正常音	1148	163	317	46
哮鸣音	169	144	18	44
爆裂音	242	25	311	25
哮鸣音和爆裂音	20	53	3	28
	正常音	哮鸣音	爆裂音	哮鸣音和爆裂音

图4