

## 基于小波变换和CNN-LSTM的肺音分类算法

张乙鹏<sup>1,2</sup>, 孙文慧<sup>1,2</sup>, 陈扶明<sup>2</sup>

1. 甘肃中医药大学信息工程学院, 甘肃 兰州 730000; 2. 中国人民解放军联勤保障部队第940医院医疗保障中心, 甘肃 兰州 730050

**【摘要】目的:**针对如何建立有助于电子听诊诊断的肺音分类模型,提出一种基于卷积神经网络(CNN)-长短期记忆网络(LSTM)的混合深度学习肺音分类模型方法。**方法:**首先使用小波变换对数据集进行特征提取,使肺音信号转化为能量熵、峰值等特征;在此基础上构建CNN和LSTM的混合算法分类模型,其中将小波变换提取的特征先输入CNN模块,能够获得数据的空间维度特征,再通过LSTM模块获得数据的时间维度特征,融合两类特征,通过模型可以将肺音分类,从而达到辅助判断患者的肺部疾病。**结果:**CNN-LSTM混合模型准确率、F1分数均明显高于其他单一模型,可达到0.948和0.950。**结论:**提出的CNN-LSTM混合模型分类准确率更高,在智能听诊领域具有广泛的潜在应用价值。

**【关键词】**肺音分类;小波变换;卷积神经网络;长短期记忆网络

**【中图分类号】**R318;TP912.35

**【文献标志码】**A

**【文章编号】**1005-202X(2024)03-0356-09

## Lung sound classification algorithm based on wavelet transform and CNN-LSTM

ZHANG Yipeng<sup>1,2</sup>, SUN Wenhui<sup>1,2</sup>, CHEN Fuming<sup>2</sup>

1. School of Information Engineering, Gansu University of Chinese Medicine, Lanzhou 730000, China; 2. Medical Security Center, The 940th Hospital of Joint Logistics Support Force of Chinese People's Liberation Army, Lanzhou Gansu, 730050, China

**Abstract: Objective** To establish a hybrid deep learning lung sound classification model based on convolutional neural network (CNN)-long short-term memory (LSTM) for electronic auscultation. **Methods** Wavelet transform was used to extract features from the dataset, transforming lung sound signals into energy entropy, peak value and other features. On this basis, a classification model based on hybrid algorithm incorporating CNN and LSTM neural network was constructed. The features extracted by wavelet transform were input into CNN module to obtain the spatial features of the data, and then the temporal features were detected through LSTM module. The fusion of the two types of features enabled the classification of lung sounds through the model, thereby assisting in the diagnosis of pulmonary diseases. **Results** The accuracy rate and F1 score of CNN-LSTM hybrid model were significantly higher than those of other single models, reaching 0.948 and 0.950. **Conclusion** The proposed CNN-LSTM hybrid model demonstrates higher accuracy and more precise classification, showcasing broad potential application value in intelligent auscultation.

**Keywords:** lung sound classification; wavelet transform; convolutional neural network; long short-term memory

### 前言

肺部疾病不仅会带来死亡的威胁,还会给患者带来自身身体机能的缺失和损伤,进一步导致患者其他伤残的发生,增加治疗费用,给原生家庭甚至社会带来负担<sup>[1]</sup>。如今空气污染日益严重,若人体吸入

大量的有毒废气或大量雾霾颗粒会导致严重的肺部疾病<sup>[2]</sup>。近几年来新冠疫情发展的严峻趋势使得智能听诊技术成为当下的一个研究热点,建立一个能够有效辅助诊断的肺音分类模型对智能听诊的发展有重要的作用,同时对临床诊断具有重要的参考意义<sup>[3]</sup>。传统的肺音分类方法基于人工特征提取和简单的机器学习算法,Falk等<sup>[4]</sup>提出心肺音信号由频谱能量的时间轨迹来表示,并且心肺音信号使用滤波器预处理。Ayari等<sup>[5]</sup>通过构建多个滤波函数,提出一种基于自适应滤波算法。Mazić等<sup>[6]</sup>构建一种简单的两层级联的SVM模式识别架构,该架构被用于记录呼吸音中所识别到的哮喘音。Chen等<sup>[7]</sup>设计一个数字听诊器以及能够检测到异常肺音的听诊系统,

**【收稿日期】**2023-12-11

**【基金项目】**国家自然科学基金(61901515,62361038);甘肃省自然科学基金(22JR5RA002)

**【作者简介】**张乙鹏,硕士,研究方向:生物医学信号检测与处理,E-mail: zyp731964329@qq.com

**【通信作者】**陈扶明,高级工程师,博士,研究方向:生物医学信号检测与处理,E-mail: cfm5762@126.com

采用K-均值算法进行特征聚类,以便识别出不同肺音。Sengupta等<sup>[8]</sup>提出一种基于传统倒谱特征统计特性的新特征集来进行计算,并使用ANN构建识别模型。张晓燕<sup>[9]</sup>提出使用小波变换和BP神经网络结合。但这些方法存在准确率低和分类结果不稳定的问题。

近年来,随着深度学习技术的成熟,肺音分类领域也进一步得到发展。深度学习技术从原始数据学习特征,深度学习技术能构建具有很强表达能力的分类模型,从而提高分类准确率和稳定性。张旭楠<sup>[10]</sup>提出基于卷积神经网络(Convolutional Neural Network, CNN)-BiGRU的心肺音分类模型,在CNN中使用一维卷积核,保留输入特征的时间维度交给BiGRU提取,充分利用CNN提取空间特征和BiGRU提取时间特征的优势,构建CNN-BiGRU模型并训练得出分类结果,但该实验方案所用肺音数据较少,且构建的模型存在特征提取不充分的问题。Choi等<sup>[11]</sup>提出一种使用注意力模块和深度学习的肺部疾病分类模型,呼吸音使用log-Mel频谱图MFCC提取。通过改进VGGish并添加一个光注意力连接模块,将有效通道注意力模块(ECA Net)应用于该模块,可以有效地对正常和5种类型的不定音进行分类,但该模型仅注意肺音信号的时间特征,从而忽略空间特征的重要性。基于此,本文提出一种基于CNN和长短期记忆网络(Long Short-Term Memory, LSTM)混合神经网络模型对肺音信号进行分类,将经过小波变换特征提取后输入到CNN网络层与LSTM网络层,充分发挥CNN提取肺音信号的空间特征和LSTM提取肺音信号的时序信息特征的能力,将提取到的特征信息输入到分类器完成分类。

## 1 CNN-LSTM 混合模型

### 1.1 小波变换特征提取

对于肺音信号而言,特征提取技术是为了从不同的信号数据中提取出能够标记不同肺音信号类型的数据而开发出来的。针对肺音信号,如何使用相应的变换方法来量化不同肺音信号在时域和频域内的特征以便于分类是极为重要的。因此,特征提取方法对于不同肺音信号的特征区分度和分类精度具有直接影响。Altan等<sup>[12]</sup>使用希尔伯特-黄变换提取统计特征,并使用深度学习算法将慢性阻塞性肺病患者和健康人群分开,但此方法准确度还有待提高。常见的声音特征参数有线性预测倒谱系数、梅尔倒谱系数等方法<sup>[13]</sup>。而本文提出采取使用小波变换提取肺音信号特征的方法。小波变换是在信号处理和

处理领域广泛应用的数学工具,它可以将信号分解成不同尺度的成分,从而帮助分析和理解信号的特征。小波变换与傅立叶变换不同,它具有时频局部性,能够更好地捕捉信号的瞬时特征。

小波变换的基本思想是将信号分解为一组小波基函数的线性组合,每个小波基函数对应不同频率和尺度的信息。这种分解允许同时获得信号的时间信息和频率信息,尤其适用于处理非平稳信号,因为小波基函数可以自适应地适应信号的不同部分。小波变换计算过程如式(1)所示:

$$W(a, b) = \int_{-\infty}^{\infty} x(t) \cdot \psi_{a,b}^*(t) dt \quad (1)$$

其中,  $W(a, b)$  是小波系数,表示信号  $x(t)$  与小波基函数  $\psi_{a,b}(t)$  的内积;  $a$  是尺度参数,控制小波基函数的宽度,通常为正值;  $b$  是平移参数,控制小波基函数在时间上的位置;  $x(t)$  是原始信号;  $\psi_{a,b}(t)$  是小波基函数,是由母小波函数进行缩放和平移得到的。不同的小波基函数有不同的数学表达式,最常见的是基于Daubechies小波、Haar小波等。小波变换提取特征的流程如图1所示。

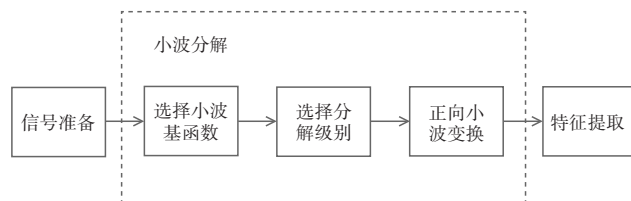


图1 小波变换特征提取流程图

Figure 1 Flowchart of feature extraction using wavelet transform

其中,信号准备是获取待处理的信号数据;小波分解的作用是为了选择合适的小波基函数和分解级别(尺度),小波分解的目的是将原始信号分解成不同尺度和频率上的小波系数子带,使用小波变换的正向变换将信号分解成一组小波系数子带;特征提取对每个小波系数子带执行以下特征提取步骤:

均值(Mean):均值衡量了信号的中心位置,是信号样本值的平均值。对于小波系数子带而言,均值表示了子带的整体振幅水平。计算过程如式(2)所示:

$$\text{Mean} = \frac{1}{N} \sum_{i=1}^N x_i \quad (2)$$

其中,  $N$  是小波系数子带的样本数量,  $x_i$  是每个样本值。

方差(Variance):方差衡量了信号的分布离散程度,是各个样本值与均值之间平均差的平方。对于小波系数子带而言,方差描述了子带的振幅分散程度。计算过程如式(3)所示:

$$\text{Variance} = \frac{1}{N} \sum_{i=1}^N (x_i - \text{Mean})^2 \quad (3)$$

标准差(Standard Deviation):标准差是方差的平方根,它表示了信号样本值与均值之间平均差的标准量度。对于小波系数子带而言,标准差是振幅的标准量度。计算过程如式(4)所示:

$$\text{Standard Deviation} = \sqrt{\text{Variance}} \quad (4)$$

能量值(Energy):能量是信号样本值的平方和,用于度量信号的强度或振幅。计算每个子带的能量可以帮助了解信号在这些不同尺度上的振幅分布。计算过程如式(5)所示:

$$\text{Energy} = \sum_{i=1}^N x_i^2 \quad (5)$$

其中, $N$ 是小波系数子带的样本数量, $x_i$ 是每个样本值。

峰值(Peak Value):峰值是信号中的最大振幅或幅值范围。峰值特征用于度量每个小波系数子带中的最大振幅,有助于了解信号在不同频率和时间尺度上的极端振幅情况。计算过程如式(6)所示:

$$\text{Peak Value} = \max(|x_i|) \quad (6)$$

峭度(Kurtosis):峭度是描述概率分布形状的统计特征,即信号中数据分布的尖锐程度,用于衡量信号的尖峰度。正峭度表示分布更尖锐或尖峭,负峭度表示分布更平坦。峭度特征用于度量每个小波系数子带中的信号分布形状。计算过程如式(7)所示:

$$\text{Kurtosis} = \frac{1}{N} \sum_{i=1}^N \left( \frac{x_i - \text{Mean}}{\text{Standard Deviation}} \right)^4 \quad (7)$$

偏度(Skewness):偏度是描述概率分布形状的统计特征,用于衡量信号分布的不对称性。正偏度表示分布向右偏斜,负偏度表示分布向左偏斜,偏度值的绝对值越大,偏斜程度越高。偏度特征用于度量每个小波系数子带中的信号分布形状。计算过程如式(8)所示:

$$\text{Skewness} = \frac{1}{N} \sum_{i=1}^N \left( \frac{x_i - \text{Mean}}{\text{Standard Deviation}} \right)^3 \quad (8)$$

能量比(Energy Ratio):能量比用于比较不同尺度或频带上的信号能量。在小波分析中,信号通过小波分解能够得到不同尺度的小波系数子带。每个子带包含不同频率和时间尺度上的信号分量。能量比可以用来比较每个子带的能量与总能量之间的比例,以了解信号在不同频率或时间尺度上的能量分布情况。

能量熵(Energy Entropy):能量熵是一种用于描述信号能量分布的统计特征。每个子带包含不同频率和时间尺度上的信号分量。能量熵可以用来衡量每个子带能量分布的不确定性或随机性。较高的能量熵表示能量分布较均匀,较低的能量熵表示能量分布较集中。计算过程如式(9)所示:

$$\text{Entropy} = - \sum_{i=1}^N p_i \log_2(p_i) \quad (9)$$

其中, $N$ 是离散能量概率分布的维度, $p_i$ 是每个能量概率分布的元素。在计算能量熵时,通常将每个小波系数子带的能量归一化,然后计算能量概率分布,并应用上述公式。

均方根(Root Mean Square, RMS):RMS是一种常用的信号特征,它表示信号振幅的均值,用于度量每个小波系数子带的整体振幅水平,可以帮助了解信号的总体振幅特性。计算过程如式(10)所示:

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad (10)$$

## 1.2 CNN

CNN是一种在图像识别领域中广泛应用的深度学习模型,它通过模拟生物学中的视觉皮层的工作原理,实现了对图像等二维数据的自动特征提取和分类,完整的二维CNN通常包括输入层、卷积层、池化层、全连接层、输出层等模块<sup>[14]</sup>。CNN的核心思想是通过卷积层、池化层和全连接层构建网络结构,以提取图像中的特征并进行分类,CNN模型结构图如图2所示。

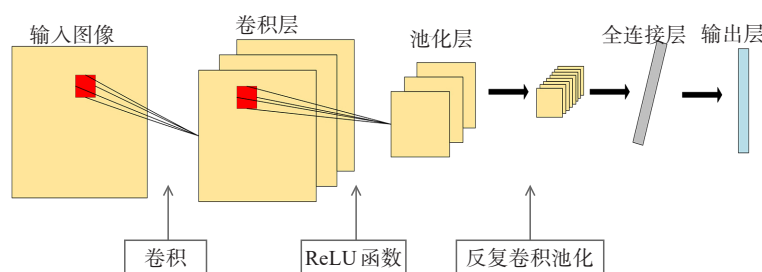


图2 CNN模型结构图

Figure 2 Structure diagram of convolutional neural network model



卷积层(Convolutional Layer)是CNN的核心部分,它由多个卷积核组成。每个卷积核在输入图像上进行滑动窗口的卷积操作,提取出图像的局部特征。卷积操作是通过将卷积核与输入图像的对应位置进行元素乘积累加得到输出特征图,通过使用不同的卷积核,卷积层可以检测出图像中的边缘、纹理等低级特征。此外,卷积层中的权重参数是通过训练数据自动学习得到的。其卷积核计算过程如式(11)所示:

$$y_j^k = f(k \sum_{i \in c_j} x_i^{k-1} * u_{ij}^k + b_j^k) \quad (11)$$

其中,\*为卷积计算, $f(x)$ 为激活函数,一般使用ReLU函数或者Sigmoid函数,作用是对卷积层的输出进行非线性映射,使得网络能够学习更复杂的特征。

卷积层后接池化层(Pooling Layer),其用于降低卷积层输出的空间维度,减少参数数量并提高计算效率,并且池化操作还具有一定的平移不变性,对输入的小的平移变化具有一定的鲁棒性。常用的池化操作是最大池化(Max Pooling),它在每个区域中选

择最大的值作为池化后的输出,从而保留重要的特征并减小图像的尺寸。

在最后一个池化层后通常接上全连接层(Fully Connected Layer),全连接层将池化层输出的特征映射与分类器相连,进行最终的分类操作。全连接层中的神经元与前一层中的所有神经元相连,每个连接都有一个权重,用于学习特征的组合。全连接层通常由一个或多个全连接隐藏层和一个输出层组成。

### 1.3 LSTM

LSTM是一种特殊的循环神经网络(Recurrent Neural Network, RNN)<sup>[15-16]</sup>,LSTM的设计旨在解决传统RNN在处理长序列时产生的梯度消失和梯度爆炸问题<sup>[17-18]</sup>,使其能够更好地捕捉长期依赖关系,既能解决短期依赖的问题又能解决长期依赖问题<sup>[19-20]</sup>。LSTM与普通RNN的主要区别在于其内部的结构,它通过使用称为“门”的结构来控制信息的流动,从而有效地处理长期依赖问题。LSTM的网络结构图如图3所示。

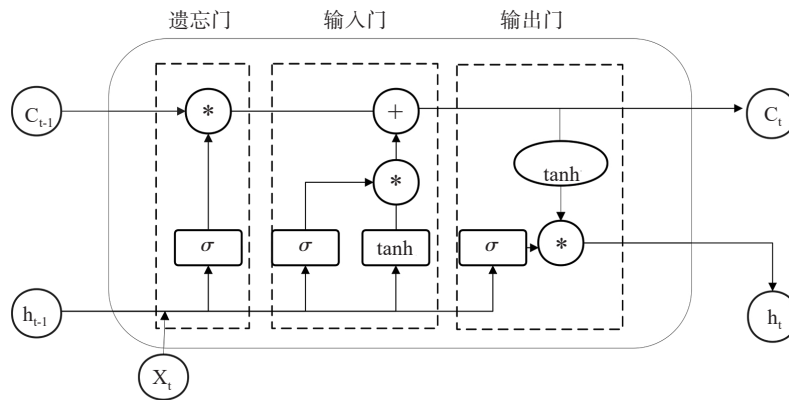


图3 LSTM网络结构图

Figure 3 Structure of long short-term memory neural network

LSTM包括3个关键组件:输入门(Input Gate)、遗忘门(Forget Gate)和输出门(Output Gate)。遗忘门决定细胞状态中哪些信息需要被遗忘,它读取上一个输出 $h_{t-1}$ 和当前一个输入 $x_t$ ,并通过一个Sigmoid激活函数( $\sigma$ )得出输出 $f_t$ ,最后与细胞状态 $C_{t-1}$ 相乘来控制哪些信息会被保留,哪些信息会被遗忘。这使得LSTM能够记住之前的相关信息,并忽略不重要的信息。计算式如式(12)所示:

$$f_t = \sigma(W_{fh}h_{t-1} + W_{fx}x_t + b_f) \quad (12)$$

输入门控制新信息的输入,它通过一个Sigmoid激活函数决定更新 $i_t$ 值,通过一个tanh函数创建新的候选值向量 $Q_t$ ,最后通过点乘与相加操作来决定多

少新信息会被添加到当前新的细胞状态 $C_t$ 中,计算式如式(13)所示:

$$\begin{aligned} i_t &= \sigma(W_{ih}h_{t-1} + W_{ix}x_t + b_i) \\ Q_t &= \tanh(W_{qh}h_{t-1} + W_{qx}x_t + b_q) \\ C_t &= f_t * C_{t-1} + i_t * Q_t \end{aligned} \quad (13)$$

输出门用于决定输出隐藏状态的哪些部分会被传递给下一个时间步。首先,它通过一个Sigmoid激活函数来控制输出的权重 $O_t$ ,然后把细胞状态通过tanh进行处理(得到一个在-1~1的值)并将它和Sigmoid门的输出相乘,最终输出确定的那部分 $C_t$ ,计算式如式(14)所示:

$$\begin{aligned} o_t &= \sigma(W_{oh}h_{t-1} + W_{ox}x_t + b_o) \\ h_t &= o_t * \tanh(C_t) \end{aligned} \quad (14)$$

1.4 CNN-LSTM混合模型肺音分类模型

在过去的研究中,大部分学者在运用CNN进行肺音分类研究时,仅考虑肺音数据的时间特性,而忽略了数据之间的空间特性,这可能会导致特征信息提取不完整。因此,本文提出一种CNN-LSTM混合模型肺音分类模型,可以在考虑数据的空间特性的

基础上考虑数据的时间特性。该模型的主要构成部分包括LSTM、CNN、全连接层和输出层,其结构如图4所示。利用CNN网络来提取数据的“空间”特性,然后采用LSTM来提取数据的“时间”特性。全连接层则根据融合后的特征来判断相应的类型。

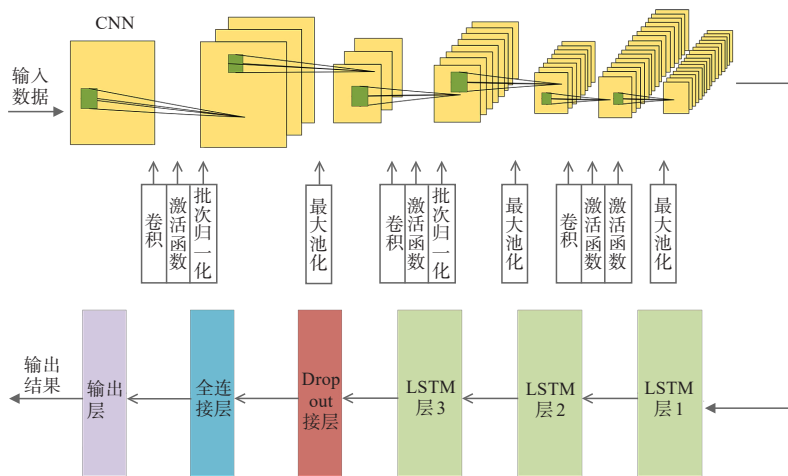


图4 CNN-LSTM混合模型结构图  
Figure 4 Structure diagram of CNN-LSTM hybrid model

为了减少训练时间并规避梯度消失和梯度爆炸问题,本模型采用反向传播算法,结合交叉熵损失函数计

算损失,并选择Adam优化器来优化模型。具体模型参数的设置如表1所示,网络训练超参数如表2所示。

表1 CNN-LSTM混合模型结构参数  
Table 1 Structural parameters of CNN-LSTM hybrid model

层	核数	尺寸	步长	隐藏节点数	输出维度	激活函数
输入层	-	-	-	-	(16, 10, 1)	-
卷积层1	32	3×3	1	-	(16, 10, 32)	-
最大池化层1	-	2×2	2	-	(8, 5, 32)	ReLU
卷积层2	64	3×3	1	-	(8, 5, 64)	-
最大池化层2	-	2×2	2	-	(4, 2, 64)	ReLU
卷积层3	128	3×3	1	-	(4, 2, 128)	-
最大池化层3	-	2×2	2	-	(2, 1, 128)	ReLU
LSTM1	-	-	-	512	(2, 256)	Tanh, Sigmoid
LSTM2	-	-	-	256	(256, 128)	Tanh, Sigmoid
LSTM3	-	-	-	128	(128, 64)	Tanh, Sigmoid
全连接层	-	-	-	-	(64, 6)	Sigmoid

基于CNN-LSTM混合模型的肺音分类模型建立主要步骤如下:(1)数据集划分与特征提取:将原始肺音数据样本随机划分为训练集和测试集,利用小波变换分别提取训练集肺音数据中信号的均值、方差、标准差、能量值、均方根等10个特征,组合成特征

向量;(2)参数设置:利用小波变换提取到的特征,通过CNN层和LSTM层进行特征学习,同时选择适宜的学习率、卷积核尺寸、LSTM单元个数和训练迭代次数等参数,以确保模型的高效学习,分别利用CNN和LSTM提取肺音数据空间特征信息和时序信息特

表2 模型训练超参数	
Table 2 Model training hyperparameters	
参数	参数值
激活函数	Sigmoid
损失函数	CrossEntropyLoss
优化器	Adam
学习率	0.001
Epoch	100
Batch size	128

征;(3)全连接层分类与参数优化:将特征向量输入全连接层,以进行分类任务,在此同时,根据验证集损失值和准确率的变化趋势,对模型参数进行微调,以进一步改进分类性能;(4)模型验证与评估:使用K折交叉验证,选择最优模型,将测试集数据输入到该模型中,根据输出结果,计算分类准确率、精确度、召回率、F1值等评价指标,并对模型进行性能评估。

## 2 结果与分析

本文实验数据处理的运行环境为 Windows11 操作系统,PyTorch 1.8.1,Python 3.8,Cuda 11.1 深度学习框架、14 vCPU Intel(R) Xeon(R) Gold 6330 CPU @ 2.00 GHz 处理器,RTX 3090(24 GB)显卡。

### 2.1 实验数据与处理

**2.1.1 实验数据集** 本次实验的数据集为 ICBHI2017 挑战赛数据集<sup>[21]</sup>,ICBHI2017 是一个用于呼吸音信号分析的公共数据集,它包含了来自不同呼吸疾病患者以及健康人的呼吸音信号,由两个不同国家的研究小组在几年内独立收集。该数据库由共 5.5 h 的记录组成,来自 920 名受试者的 126 个注释音频样本,这些录音是使用不同设备收集的,持续时间为 10~90 s,还提供获取录音的胸部位置。呼吸科医疗专家对这些肺音进行注释,其中共包含 8 种类型:哮喘(Asthma)、支气管扩张(Bronchiectasis)、细支气管炎(Bronchiolitis)、慢性阻塞性肺病(COPD)、健康(Healthy)、下呼吸道感染(LRTI)、肺炎(Pneumonia)、上呼吸道感染(URTI)。具体数量分析如图 5 所示。

**2.1.2 数据处理** 由图 5 可知,该数据集每个类别的数据量不一样,类不平衡现象较为严重,其中 2 种疾病 Asthma 和 LRTI 分别仅有 1 个和 2 个,因此首先去除这两种疾病类型之后再进一步的进行数据预处理,最终数据集共分为 6 类 917 条数据。该数据集数据量较小并且类不平衡现象较为明显,不利于深度学习的训练,因此对数据进行预处理的手段包括数据增强和采样。

数据增强增加训练数据的多样性、减少过拟合。

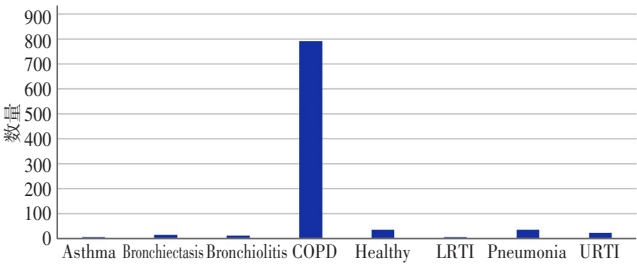


图5 原始数据统计图  
Figure 5 Raw data statistics

针对本实验的音频信号,采用的数据增强手段有:(1)时间拓展:在原始肺音信号中添加静音或重复部分来增加时间长度;(2)频率扩展:通过改变语音信号的音高或音调来增加频率变化;(3)背景噪声:在肺音信号中添加背景噪声,模拟真实场景中的环境噪声,生成一个白噪声信号,通过创建的一个一阶低通 Butterworth 滤波器将噪声信号添加到肺音信号中;(4)数据剪辑:从原始肺音信号中随机剪辑一段音频,以产生多个不同长度的肺音片段。

对于类不平衡的问题采取过采样和欠采样的方式来调节数据集类的分布,过采样是指对于数量少的类在进行数据增强时多进行几次随机增强,欠采样是指对于数据量多的类对增强后的数据进行随机抽样采样,最终控制采样后的数据类更均匀。图 6 展示了预处理后数据分布情况,可以得出原始数据类分布严重不平衡问题,经过数据增强后有较大改善。

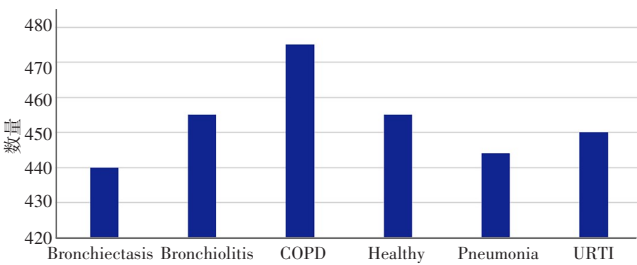


图6 处理后数据统计图  
Figure 6 Data statistics after processing

### 2.2 K折交叉验证选择最优模型

K折交叉验证是深度学习中一种非常有用的技术,用于选择深度学习模型。在深度学习中,通常需要评估一个模型的性能,以确定它是否能够有效地泛化到未见过的数据。常见的方法是将数据集划分为训练集、验证集和测试集,其中用于训练模型的是训练集;验证集用来验证模型的性能,能够及时调整模型超参数;测试集用来评价模型的性能。然而,这种方法存在随机性的问题。如果只使用一次数据划分来训练模型,那

么模型结果可能会受到特定划分的影响。而K折交叉验证可以帮助更好地理解模型在不同数据子集上的性能,减少由于数据划分的随机性引起的评估偏差,从而提高模型评估的可靠性。

首先,将数据集中的训练集分为K个折叠,每个折叠包含大致相等数量的样本。这个划分是随机的,确保每个折叠都是独立的,没有重叠。选择一个折叠作为验证集,将其余4个折叠合并成一个训练集。在训练集上训练分类模型,使用验证集来评估模型的性能。完成5次迭代后,会得到5个性能指标,每个指标对应一个验证集上的模型性能评估。选取最优的模型,使用测试集进行测试验证。本文采取K折交叉验证进行最优模型选择,K取值为5,表3为3种模型输出得到的验证集上的准确率(Acc)和损失率(Loss)。

表3 各模型在验证集上的准确率与损失率  
Table 3 Accuracy and loss rate of each model on the validation set

模型	折数	准确率	损失率
CNN 模型	1	0.952	0.219
	2	0.955	0.258
	3	0.954	0.249
	4	0.954	0.261
	5	0.961	0.316
LSTM 模型	1	0.943	0.301
	2	0.941	0.263
	3	0.939	0.285
	4	0.933	0.335
	5	0.948	0.304
CNN-LSTM 混合模型	1	0.965	0.399
	2	0.959	0.410
	3	0.959	0.400
	4	0.969	0.355
	5	0.967	0.280

对于表3的结果,通过综合考虑准确率和损失率来选择最优模型,选取CNN的第2组、LSTM的第5组、CNN-LSTM混合模型的第5组作为最佳模型。从图7可以看出,所有模型都能快速收敛并趋于稳定,但通过对准确率和损失率的综合考虑,CNN-LSTM混合模型在数据集上表现出最佳性能。

2.3 特征可视化

为了测试模型的性能,在使用相同测试集的前提下,使用K折交叉验证选择的CNN-LSTM混合模型、CNN模型和LSTM模型进行对比。当学习率是0.001时,为了对比该模型特征提取的性能,分别获取CNN、LSTM、

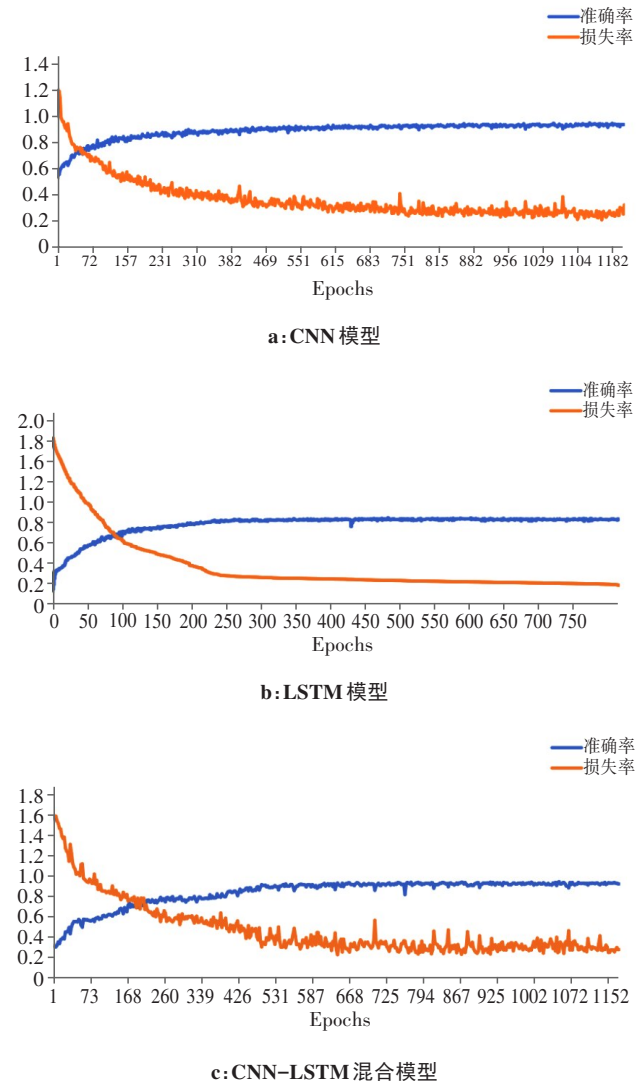


图7 模型准确率和损失率曲线图  
Figure 7 Model accuracy and loss rate curves

CNN-LSTM混合模型所提取的特征信息,利用t-SNE技术对特征提取进行可视化显示,具体结果如图8所示。图8中显示CNN模型提取的特征在空间上分布较为混杂,特征之间的交叉较多。这表明CNN模型的特征提取能力相对较弱,不能够明确区分不同的类别,从而导致分类错误;虽然LSTM模型的特征边界相对明显,但还是存在类分散和类交叉的情况;与前两种模型相比,本文提出的CNN-LSTM混合模型在t-SNE图中显示出更明显的特征边界,说明该模型能够更好地区分不同类别,并且提取的特征更容易区分,提高分类的准确性。CNN-LSTM混合模型提取的特征更具区分度,可以在实际应用中更好地处理类别不平衡或相似性高的情况。

2.4 评价指标

为了全面评估网络模型的性能,首先采用准确率作为评价指标,准确率是一种便捷的指标,可以直观地对比不同模型的性能;其次选择使用混淆矩阵来详细展示网络模型对各个类别的预测情



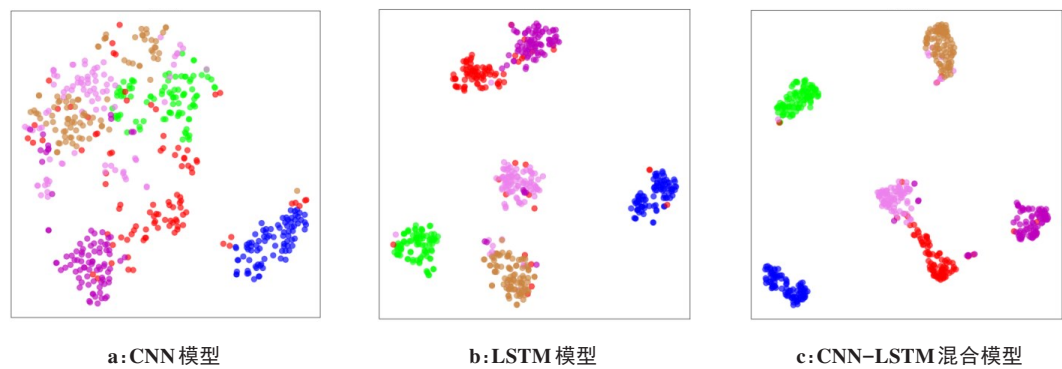


图8 特征可视化  
Figure 8 Feature visualization

况<sup>[22]</sup>。混淆矩阵综合记录了模型预测结果与真实值之间的差异,若以病情分为正常和异常二分类为例,TP代表模型预测为正常,且真实类型为正常;FP代表模型预测为正常,但真实类型为异常;FN代表模型预测为异常,但真实类型为正常;TN代表模型预测为异常,且真实类型为异常。

基于混淆矩阵结果,可延伸出二级指标,包括精确度(Precision)、召回率(Recall)。对预测值而言,在被预测为正常的样本中,预测正确的概率即为精确度,其计算过程如式(15)所示:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

(15)

精确度是对预测值而言,在被预测为正常的样本中,预测正确的概率,其计算过程如式(16)所示:

$$Precision = \frac{TP}{TP + FP}$$

(16)

召回率是对真实值而言,预测为正常占样本中实际正常的比例,其计算过程如式(17)所示:

$$Recall = \frac{TP}{TP + FN}$$

(17)

在实际的肺音分类过程中,面临多种疾病的存在。因此,在利用网络模型进行分类时,必须确保分类结果既不遗漏也不重复,这使得召回率成为一个关键的评价指标。此外,为了避免出现评价指标之间的冲突和难以综合评估的情况,本文引入F1分数作为评估指标。F1分数综合考虑了网络模型的召回率和特异度。F1分数的取值为0~1,分数越高表示网络模型在不同样本的识别能力越强,公式如下所示:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

(18)

2.5 性能评估

使用测试集分别对3种模型进行测试,其分类结果的混淆矩阵如图9所示,其中本文所提的CNN-LSTM混合模型模型准确率最高,可达到0.948。分别计算各模型的召回率、精确度以及F1分数,见表4。

在相同的测试集下,与其他两种算法相比,CNN-LSTM混合模型模型表现出较好的性能:(1)CNN-LSTM混合模型模型的准确率达到0.948,远高于另外两种单一模型,这表明该模型在对肺音数据进行分类时能够非常

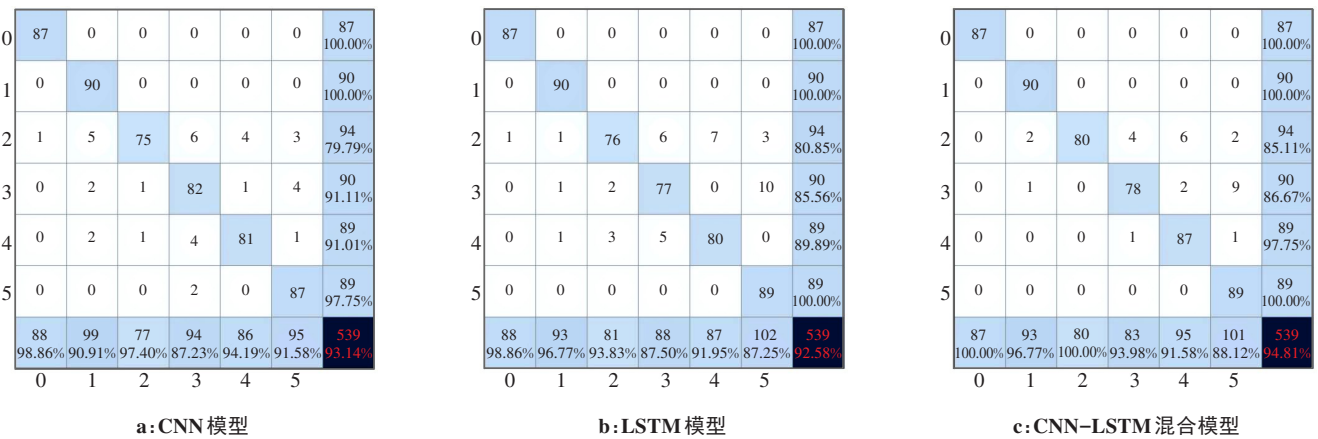


图9 各模型混淆矩阵  
Figure 9 Confusion matrix of each model



表4 各模型评估指标  
Table 4 Evaluation indicators for each model

模型	准确率	召回率	精确度	F1 分数
CNN 模型	0.931	0.934	0.933	0.933
LSTM 模型	0.926	0.927	0.927	0.927
CNN-LSTM 混合模型	0.948	0.951	0.949	0.950

准确地判断样本的类别;(2)该模型的召回率显著高于单独的CNN和LSTM模型,达到0.951,高召回率意味着模型能够有效地捕捉到正样本,减少漏报的情况,这对于肺音分类非常重要;(3)F1分数达到0.950,进一步证明了该模型在分类不同类别方面的能力,F1分数是综合考虑精确度和召回率的指标,高F1分数值表明模型在样本的分类上表现出优异的性能。

2.6 对比实验

Bardou等<sup>[23]</sup>论文中,Bardou\_CNN模型算法的分类识别准确率最高,为0.934。为了说明Bardou\_CNN模型和本文CNN-LSTM模型的优劣性,将用Bardou\_CNN模型替代本文算法下的CNN-LSTM模型。经实验后,Bardou\_CNN的分类识别准确率为0.896。在本文同一数据下,Bardou\_CNN的分类识别准确率比CNN-LSTM算法的分类识别准确率(0.948)低0.052,说明本文提出的CNN-LSTM模型优于Bardou\_CNN。

3 结 语

本文针对肺音信号分类问题,提出一种基于CNN-LSTM混合深度学习分类模型。首先,为了应对数据集分类不均匀的问题,采用数据增强技术,这有效地减轻了模型训练中的过拟合和数据不平衡问题;其次,利用小波变换技术,提取肺音信号相关特征;最后,在数据输入后,研究充分利用CNN和LSTM的优势,CNN用于提取空间特征,LSTM则用于提取时序特征,通过调整参数并使用K折交叉验证,选择最优模型。最终的测试结果表明,该模型在准确率和F1分数方面表现出卓越的性能,分别达到0.948和0.950,相比CNN模型、LSTM模型,本文提出的CNN-LSTM混合模型分类准确率更高,在智能听诊领域具有广泛的潜在应用价值。

【参考文献】

[1] 胡建平,饶克勤,钱军程,等.中国慢性非传染性疾病经济负担研究[J].中国慢性病预防与控制,2007,15(3):189-193.  
Hu JP, Rao KQ, Qian JC, et al. The study of economic burden of chronic non-communicable diseases in China[J]. Chinese Journal of Prevention and Control of Chronic Non-Communicable Diseases, 2007, 15(3): 189-193.  
[2] Hollman A. An ear to the chest: an illustrated history of the evolution of the stethoscope[J]. J R Soc Med, 2002, 95(12): 626-627.  
[3] 敖一鹭.便携式电子心肺音听诊器的研制[D].重庆:重庆大学,2016.  
Ao YL. Research on portable electronic heart sounds and lung sounds

stethoscope[D]. Chongqing: Chongqing University, 2016.  
[4] Falk TH, Chan WY. Modulation filtering for heart and lung sound separation from breath sound recordings[J]. Annu Int Conf IEEE Eng Med Biol Soc, 2008, 2008: 1859-1862.  
[5] Ayari F, Ksouri M, Alouani AT. Computer based analysis for heart and lung signals separation [C]//2013 International Conference on Computer Medical Applications (ICCMA). Piscataway, NJ, USA: IEEE, 2013: 1-6.  
[6] Mazić I, Bonković M, Džaja B. Two-level coarse-to-fine classification algorithm for asthma wheezing recognition in children's respiratory sounds[J]. Biomed Signal Process Control, 2015, 21: 105-118.  
[7] Chen CH, Huang WT, Tan TH, et al. Using K-nearest neighbor classification to diagnose abnormal lung sounds[J]. Sensors (Basel), 2015, 15(6): 13132-13158.  
[8] Sengupta N, Sahidullah M, Saha G. Lung sound classification using cepstral-based statistical features[J]. Comput Biol Med, 2016, 75: 118-129.  
[9] 张晓燕.基于BP神经网络的肺音识别与诊断研究[J].电子测试,2016(13):111-113.  
Zhang XY. Research of lung sound recognition and diagnosis based on BP neural network[J]. Electronic Test, 2016(13): 111-113.  
[10] 张旭楠.基于深度学习的听诊信号识别系统的研究与设计[D].镇江:江苏大学,2022.  
Zhang XN. Research and design of auscultation signal recognition system based on deep learning[D]. Zhenjiang: Jiangsu University, 2022.  
[11] Choi Y, Lee H. Interpretation of lung disease classification with light attention connected module[J]. Biomed Signal Process Control, 2023, 84: 104695.  
[12] Altan G, Kutlu Y, Allahverdi N. Deep learning on computerized analysis of chronic obstructive pulmonary disease[J]. IEEE J Biomed Health Inform, 2020, 24(5): 1344-1350.  
[13] 陈树,于海波.一种改进的特征提取方法在语音识别中的应用[J].传感器与微系统,2018,37(5):154-157.  
Chen S, Yu HB. Application of an improved feature extraction method in speech recognition[J]. Transducer and Microsystem Technologies, 2018, 37(5): 154-157.  
[14] Ketkar N, Moolayil J. Deep learning with Python: learn best practices of deep learning models with PyTorch[M]. 2nd ed. Berkeley, CA, USA: Apress, 2021: 197-242.  
[15] 杨丽,吴雨茜,王俊丽,等.循环神经网络研究综述[J].计算机应用,2018,38(S2):1-6.  
Yang L, Wu YQ, Wang JL, et al. Research on recurrent neural network [J]. Journal of Computer Applications, 2018, 38(S2): 1-6.  
[16] 胡凯,郑翥,卢飞宇,等.基于深度学习的行为识别算法综述[J].南京信息工程大学学报(自然科学版),2021,13(6):730-743.  
Hu K, Zheng F, Lu FY, et al. A survey of action recognition algorithms based on deep learning [J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition), 2021, 13(6): 730-743.  
[17] 董立红,肖纯朗,叶鸥,等.一种基于CAEs-LSTM融合模型的窃电检测方法[J].电力系统保护与控制,2022,50(21):118-127.  
Dong LH, Xiao CL, Ye O, et al. Electricity theft detection method based on a CAEs-LSTM fusion model[J]. Power System Protection and Control, 2022, 50(21): 118-127.  
[18] 李滨,王靖德,梁水莹,等.基于长短期记忆循环神经网络的AGC实时控制策略[J].电力自动化设备,2022,42(3):128-134.  
Li B, Wang JD, Liang SY, et al. AGC real-time control strategy based on LSTM recurrent neural network[J]. Electric Power Automation Equipment, 2022, 42(3): 128-134.  
[19] 黄曼云,王天昊,卫志农,等.基于长短期记忆网络的UKF动态谐波状态估计[J].电力系统保护与控制,2022,50(11):1-11.  
Huang MY, Wang TH, Wei ZN, et al. Dynamic harmonic state estimation of an unscented Kalman filter based on long short-term memory neural networks[J]. Power System Protection and Control, 2022, 50(11): 1-11.  
[20] 毛煜,尚海昆,于卓琦.基于长短期记忆网络的电网同调机群快速辨识[J].电气工程学报,2022,17(2):201-207.  
Mao Y, Shang HK, Yu ZQ. A fast prediction method of coherent generators based on long short-term memory network[J]. Journal of Electrical Engineering, 2022, 17(2): 201-207.  
[21] Rocha BM, Filos D, Mendes L, et al. An open access database for the evaluation of respiratory sound classification algorithms [J]. Physiological Measurement, 2019, 40(3): 035001.  
[22] 陈德昊,林建恒,衣雪娟,等.基于小波包时频图特征和卷积神经网络的水声信号分类[J].声学技术,2021,40(3):336-340.  
Chen DH, Lin JH, Yi XJ, et al. Classification of underwater acoustic signals based on time-frequency map features of wavelet packet and convolutional neural network[J]. Technical Acoustics, 2021, 40(3): 336-340.  
[23] Bardou D, Zhang K, Ahmad SM. Lung sounds classification using convolutional neural networks[J]. Artif Intell Med, 2018, 88: 58-69.

(编辑:陈丽霞)