

分类号: TP391

单位代码: 11407

密 级: 公开

学 号: 20207213



# 北方民族大学

硕士学位论文

## 基于 CNN-Transformer 的肺音信号分类研究

Classification of lung sound signals based on CNN-Transformer

学 位 申 请 人: 田思远

指 导 教 师: 韦海成

申 请 的 学 位 门 类: 工学硕士

专 业 名 称: 电路与系统

研 究 方 向: 生物医学电信号分析与处理

所 在 学 院: 电气信息工程学院

论 文 完 成 日 期: 2023年04月

## 摘要

近年来，慢性阻塞性肺疾病(Chronic Obstructive Pulmonary Disease, COPD)成为人类主要的死亡原因之一，严重威胁人类的生命安全。但是由于该疾病早期症状并不明显，往往延误患者治疗的最佳时机。因此，对 COPD 患者进行早期诊治非常重要。本文利用深度学习技术对肺音信号进行分类，有助于 COPD 的早期预防和筛查。本文主要研究内容如下：

(1) 针对肺音信号中存在噪声的问题，采用改进的变分模态分解(Improved Variational Mode Decomposition, IVMD)算法对肺音信号进行去噪。首先搭建了肺音信号采集系统，其次利用 IVMD 算法对 ICBHI 公有数据集的 500 组 COPD 患者和实验室自采的 500 组健康人肺音信号进行去噪，最后根据相关系数筛选固有模态分量(Intrinsic Mode Function, IMF)进行重构肺音信号，得到纯净的肺音信号。

(2) 针对单一肺音信号特征不能充分地表征肺音信号的问题，采用融合特征参数，充分地提取了肺音信号的特征。根据肺音信号的时间序列特性和非线性非平稳特性，分别提取了线性预测倒谱系数(Linear Predictive Cepstral Coefficient, LPCC)和梅尔频率倒谱系数(Mel-Frequency Cepstral Coefficient, MFCC)，再将上述两种特征参数进行融合，得到包含更多肺音信号信息的融合特征，为后续的分类奠定基础。

(3) 针对传统肺音信号分类方法准确率低、常用的深度学习模型层数高、泛化性不强、无法提取长时间序列等问题，采用基于 CNN-Transformer 的肺音信号分类模型。利用卷积神经网络模型自动提取肺音信号特征的特点，搭建了 CNN 网络模型，对肺音信号进行分类；再利用支持向量机(Support Vector Machines, SVM)替换原始网络模型中的 Softmax 函数，搭建了 CNN-SVM 网络模型，提升网络模型的泛化性；最后结合具有并行结构的 Transformer，提取肺音信号的长时间序列，采用 CNN-Transformer 网络模型，提升分类的准确率。

实验结果表明，IVMD 算法可以显著去除噪声，提升肺音信号的信噪比。本文所采用的 CNN-Transformer 网络模型对肺音信号进行分类时，准确率、灵敏度、精确度和 F1-score 分别为 95.70%、94.62%、93.87%、93.88%，比 CNN-SVM 网络模型分别提升了 7.32%、2.49%、5.22%、1.51%，为 COPD 患者进行早期筛查提供可能性。

**关键词：**肺音信号分类；特征融合；变分模态分解；卷积神经网络

## Abstract

In recent years, Chronic Obstructive Pulmonary Disease (COPD) has become one of the main causes of death in human beings, which seriously threatens the life safety of human beings. But because the early symptoms of the disease are not obvious, the best time to treat patients is often delayed. Therefore, it is important to diagnose and treat COPD patients at an early stage. In this thesis, deep learning technology is used to classify lung sound signals, which can help in the early prevention and screening of COPD. The main research content of this thesis is as follows

(1) To address the problem of noise in lung sound signals, the Improved Variational Mode Decomposition (IVMD) algorithm was used to denoise the lung sound signals. Firstly, a lung sound signal acquisition system was built. Secondly, the IVMD algorithm was used to denoise the lung sound signals of 500 groups of COPD patients from the ICBHI public dataset and 500 groups of healthy people from the laboratory. Finally, according to the correlation coefficients, the Intrinsic Mode Functions (IMF) were selected to reconstructed the lung sound signals and the pure lung sound signals were obtained.

(2) To address the problem that a single lung sound signal cannot adequately characterize the lung sound signal, the fusion feature parameters were used to fully extract the features of the lung sound signal. The Linear Predictive Cepstral Coefficient (LPCC) and Mel-Frequency Cepstral Coefficient (MFCC) were extracted according to the time series characteristics and nonlinear non-smooth characteristics of the lung sound signal. Then the above two feature parameters was used to obtain fusion features containing more lung sound signal information, which laid the foundation for the subsequent classification.

(3) A CNN-Transformer-based lung sound signal classification model was proposed to address the problems of low accuracy of traditional lung sound classification, high layer count of commonly used deep learning models, poor generalization, and inability to extract long time sequences. The CNN network model was built to classify the lung sound signal by using the feature of convolutional neural network model to automatically extract the features of lung sound signal; then the Softmax function in the original network model was

replaced by Support Vector Machines (SVM) to build a CNN-SVM network model to improve the generalization of the network model. The CNN-Transformer network model was proposed to improve the accuracy of classification by combining the Transformer with parallel structure and extracting the long time sequence of lung sound signal.

The experimental results show that the IVMD algorithm can significantly remove the noise and improve the signal-to-noise ratio of the lung sound signal. The accuracy, sensitivity, precision and F1-score of the proposed CNN-Transformer network model for lung sound signal classification were 95.70%, 94.62%, 93.87% and 93.88%, respectively, which improved 7.32%, 2.49%, 5.22% and 1.51% better than the CNN-SVM network model, providing COPD patients with the possibility of early screening.

**Key words :** Lung Sound Signal Classification; Feature Fusion; Variational Modal Decomposition; Convolutional Neural Network

# 目 录

第一章 绪论.....	1
1.1 研究背景及意义.....	1
1.2 国内外研究现状.....	1
1.2.1 肺音信号预处理的研究现状.....	2
1.2.2 肺音信号特征提取的研究现状.....	3
1.2.3 肺音信号分类的研究现状.....	4
1.3 本文研究内容.....	6
1.4 本文结构安排.....	7
第二章 肺音信号预处理及分类的理论基础 .....	8
2.1 肺音信号的理论基础.....	8
2.1.1 肺音的基本概念及分类.....	8
2.1.2 肺音信号的特性.....	8
2.2 肺音信号预处理的理论基础 .....	9
2.2.1 经验模态分解.....	9
2.2.2 变分模态分解.....	10
2.3 基于深度学习的肺音信号分类模型理论基础 .....	14
2.3.1 卷积神经网络的理论基础.....	14
2.3.2 Transformer 网络模型的理论基础.....	16
2.4 本章小结.....	18
第三章 肺音信号的采集及预处理 .....	19
3.1 肺音信号的采集.....	19
3.2 肺音信号的预处理.....	20
3.3 基于 EMD 的肺音信号去噪研究.....	22
3.4 基于 IVMD 的肺音信号去噪研究.....	24
3.5 实验结果与分析.....	27
3.5.1 去噪评价指标.....	27
3.5.2 实验结果对比.....	28
3.6 本章小结.....	30
第四章 肺音信号的特征提取 .....	31
4.1 肺音信号的线性预测倒谱系数 .....	31
4.1.1 线性预测系数.....	31
4.1.2 线性预测倒谱系数.....	33
4.2 肺音信号的梅尔频率倒谱系数提取 .....	33
4.2.1 梅尔频率倒谱系数.....	33

4.2.2 梅尔频率倒谱系数提取.....	35
4.3 肺音信号的 LPCC-MFCC 特征参数融合 .....	36
4.4 本章小结.....	37
第五章 基于 CNN-Transformer 的肺音信号分类模型研究 .....	38
5.1 基于 CNN 的肺音信号分类模型研究 .....	38
5.2 基于 CNN-SVM 肺音信号分类模型研究.....	39
5.2.1 支持向量机.....	39
5.2.2 基于 CNN-SVM 的肺音信号分类模型 .....	40
5.3 基于 CNN-Transformer 的肺音信号分类模型研究.....	41
5.4 实验结果与分析.....	42
5.4.1 实验数据及参数设置.....	42
5.4.2 评价指标.....	42
5.4.3 分类结果与分析.....	43
5.5 本章小结.....	46
第六章 总结与展望 .....	47
6.1 总结.....	47
6.2 展望.....	47
参考文献.....	49

## 第一章 绪论

### 1.1 研究背景及意义

2017 年的全球疾病负担研究报告指出，2017 年全球因吸烟、大气污染和职业污染等原因导致 391 万人患慢性呼吸系统疾病死亡，占全球死亡人数的 7%，其中慢性阻塞性肺病是中国死亡人数的第三大原因<sup>[1]</sup>。世界卫生组织 2021 年发布的全球癌症数据显示，2020 年中国新发癌症 457 万人，占全球新发癌症人数的 23.7%，其中肺癌新发人数占全部癌症新增人数的 17.9%<sup>[2]</sup>。但是此类慢性呼吸系统疾病早期症状不明显，患者容易错过最佳治疗时间，导致病情加重，对人体造成不可逆的危害。由此可见，为了保护人类的生命安全，采取有效手段对慢性呼吸系统疾病进行早期预防并及时治疗尤为重要。

肺音听诊因其经济、操作简单、非侵入式的特点，成为常用的临床肺部疾病诊断方式之一<sup>[3]</sup>。但是目前常用的听诊方式存在不足，比如数据不能长期存储，听诊结果受医生的临床听诊经验和技能娴熟程度等主观因素的影响。随着现代科技的进步，肺功能检查、X 射线诊断、胸透、肺动脉血气相等精确的诊断方法成为临床呼吸系统疾病诊断方式。但是上述方法操作复杂、就诊费用高昂且含有辐射危害，不利于呼吸系统疾病的日常筛查。因此需要研究一种便捷、操作简单的呼吸系统疾病筛查方法。

得益于现代数字信号处理技术的迅速发展，越来越多的学者开始对肺音信号的分类问题进行研究。肺音信号分类主要分为三个步骤：预处理、特征提取、识别分类，通常有傅里叶变换、短时傅里叶变换(STFT)、小波变换等传统方法，支持向量机(SVM)和隐马尔可夫模型等机器方法，人工神经网络、卷积神经网络和循环神经网络(RNN)等深度学习方法。然而，在实际研究中存在一些不足，比如未对肺音信号进行预处理，未将原始信号的设备噪声、环境噪声和心音信号去除，导致检测精度和效率较低。

本文将对肺音信号进行采集并存储，然后对肺音信号进行降采样，再经过去噪得到纯净的肺音信号，根据肺音信号的特性进行特征提取与融合，充分保留肺音信号的有效特征，最后利用深度学习技术对融合特征进行分类。该方法不仅有利于监督肺部健康状况，降低就诊难度，提高就诊效率，而且对医生进一步诊断和治疗具有重要意义。

### 1.2 国内外研究现状

早期国内外研究学者利用肺音信号的时长、相位、幅度等客观特征对肺音信号进行分析<sup>[4]</sup>，

但因肺音信号的非线性非平稳特性，单一的时域特征不能充分反映其本质。于是，人们逐渐采用频域分析方法和统计学特征<sup>[5]</sup>对肺音信号进行分析，并将数字信号处理同深度学习等人工智能结合，对肺音信号进行深入研究。

### 1.2.1 肺音信号预处理的研究现状

肺音信号预处理包括对肺音信号进行分帧、加窗、去噪等步骤。因为在肺音信号的采集过程中会混入环境噪声以及心音信号，电子听诊器虽然可以降低环境噪声对肺音信号的影响，但是去噪效果不佳。这将影响特征提取和肺音信号分类的准确性，因此对肺音信号进行去噪至关重要。

近些年，国内外的研究人员已经进行了大量的尝试，提出了许多肺音信号预处理的方法，如小波变换、非负矩阵分解和自适应分解滤波等方法。

因为小波变换在高频处具有较高时间分辨率，在低频处具有较高频率分辨率，能适应时频信号分析的要求，从而可聚焦到信号的任意细节，所以小波变换在肺音信号去噪中得到了广泛应用。石陆魁等人对肺音信号进行五层小波变换，达到滤除高频噪声的目的<sup>[6]</sup>。因为肺音信号频率和心音信号频率具有高度重叠，使用单一小波变换去噪对心音信号干扰的滤除效果不理想。姚小静等人先对原始信号进行高通去噪，然后采用软阈值法对小波变换后的每一层细节信息进一步处理，滤除肺音信号中的心音干扰<sup>[7]</sup>。Pouyani 等人采用 Db14 作为离散小波变换的母小波对原始信号进行 8 层小波变换，去除肺音信号中的部分噪声<sup>[8]</sup>。Tripathy 等人先对肺音信号进行分帧处理，再用有固定边界点的经验小波变换对分帧后的信号进行分解<sup>[9]</sup>，为后续提取特征得到纯净的肺音信号。但是不同的小波基分析结果可能不同，小波变换的分解层数和小波系数阈值等参数的选择对分析结果也存在影响，因此选择合适的小波基比较困难。

现实采集的肺音信号通常是单通道的，非负矩阵分解因其实现简便、占用存储空间少、适用于单通道信号分解的特点，所以已经成为肺音信号预处理的主要方法之一。郭鹿鸣先对原始信号进行短时傅里叶变换，将信号变换到时频域上，然后用非负矩阵分解进行降维处理得到原始信号分量集，再进行无监督自学习的聚类，将肺音信号和心音信号分离<sup>[10]</sup>。但是传统的基于非负矩阵分解的心肺音分离方法未参考信号中心音和肺音的标签信息，所以分离效果受限。朱俊霖等人采用带标签约束的非负矩阵分解，将心音和肺音频率的特点作为约束条件，得到更好的分离效果<sup>[11]</sup>。雷志彬将纯净的心音信号和肺音信号混合后，再进行短时傅里叶变换得到时频谱，然后利用回归迁移非负矩阵分解从混合信号时频谱中得到心音和肺音的时频谱，最后通过逆短时傅里叶变换重构出心音和肺音的时域信号，实验表明分离效果可观<sup>[12]</sup>。Muñoz-Montoro 等人搭建一个基于非负矩阵分解的并行系统，将非负矩阵分解、聚类与软掩蔽去噪相结合来分解单通道混合肺音信号，

有效提升肺音信号的质量<sup>[13]</sup>。Grooby 等人提出了一种基于非负矩阵共因子分解的方法，根据加权因子实现心音、肺音和噪声分离，得到高质量的肺音信号<sup>[14]</sup>。但是不同的数据集可能需要不同的非负矩阵分解基数，因此非负矩阵分解的基数难以确定，缺乏自适应性。

为了更好地对非线性非平稳的肺音信号进行预处理，自适应分解方法应运而生。因为无需先验知识，所以在肺音信号的去噪方向已经得到广泛应用。Aziz 等人利用经验模态分解(Empirical Mode Decomposition, EMD)对正常肺音信号和肺炎信号进行去噪处理，根据肺音频率范围将得到的固有模态分量重构，主要消除了高频噪声<sup>[15]</sup>。Khan 等人对非慢性肺疾病信号和 COPD 信号进行经验模态分解去噪，以固有模态分量的能量为依据重构信号，实验表明此方法得到的信号满足特征提取和分类的需要<sup>[16]</sup>。高鸣蕾对 RPSEMD 进行改进，再结合小波软阈值，对肺音信号进行去噪，比单纯使用经验模态分解去噪效果有所提升<sup>[17]</sup>。刘毅提出一种最优有理四次 Hermite 包络线插值算法，首先利用经验模态分解自适应地分解肺音信号，再对重构信号进行经验模态分解和单通道盲源分离，去除其他噪声并抑制心音信号干扰，从而得到相对纯净的肺音信号<sup>[18]</sup>。但是传统的经验模态分解算法存在模态混叠、端点效应等问题，这些都会对肺音信号分离产生影响。本文首先对肺音信号进行降采样和归一化处理，再使用 EMD 系列分解方法和改进的变分模态分解对肺音信号进行去噪，提升肺音信号的质量，更适用于后续的工作。

### 1.2.2 肺音信号特征提取的研究现状

肺音信号特征提取主要是去除与分类无关的信息，以找到最有效的信号特征来构成用于分类的特征向量。近年来，常用的肺音信号特征提取方法主要有时域、频域、时频域联合等方法。

在前期的肺音研究中，由于技术条件和分类方法受限以及研究的重点是肺音信号采集设备的研制，时域特征提取因其操作简单且直观成为一种主要的特征提取方法。主要有短时能量、短时过零率、呼气相和吸气相的持续时间之比、幅值、包络等<sup>[19]</sup>。崔星星等人提出一种肺音信号特征提取方法，首先提取肺音信号的短时能量和短时过零率，然后构建特征向量实现特征数字化，为后续肺音信号的识别和诊断提供理论支撑<sup>[20]</sup>。黄晓旗等人从医学角度出发，将呼气相全肺平均肺密度、用力肺活量(FVC)作为特征对 COPD 患者进行识别<sup>[21]</sup>。陈仕锋等人自采数据，根据用力肺活量占预计值百分比率(FVC%)对首次确诊慢阻肺患者进行分析，划分慢阻肺患者的严重程度<sup>[22]</sup>。Infante 等人提取肺音信号的峰度、方差等时域特征作为分类的依据<sup>[23]</sup>。Rahman 等人提取肺音信号的标准差、偏度、峰度、过零率等时域特征，对健康人、COPD 患者和哮喘患者进行分类<sup>[24]</sup>。但是用于判断呼吸气相的设备复杂、条件严苛，时域特征并不能充分反映肺音信号非线性非平稳的特点。为了更充分地提取肺音信号的特征，研究人员开始从频域对肺音信号进行特征提取。

Abougabal 等人提出一种对儿童肺音信号进行特征提取的方法，对去噪后的重构信号进行离散小波变换得到离散小波系数作为特征，进行下一步聚类分析<sup>[25]</sup>。Islam 等人首先将对信号进行幅度归一化，去除受试者的生理、年龄和性别等因素的影响，然后进行小波变换，从 d2, d3, d4, d5 和 d6 子频段提取统计特征，使用韦尔奇的方法从每个肺音周期估计功率谱密度，然后将其分解为均匀的子带，从每个子带中计算一组统计特征，以对正常和哮喘受试者进行分类<sup>[26]</sup>。Göögüs 等人先将肺音信号分为吸气和呼气相，然后使用 FFT、自回归模型和自回归滑动平均模型从吸气和呼气信号中分别提取功率谱密度，并对功率谱密度进行离散小波变换，选取前五个频带提取相邻频率区间的平均值、最大值、标准差和平均值之比等四个统计特征，用来区分严重哮喘、中度哮喘、轻度哮喘、COPD、过敏性肺炎和正常肺音<sup>[27]</sup>。

随着研究学者对肺音信号的深入研究，能够同时从时间和频率两个方面表征信号，使得时频域特征联合逐渐成为特征提取研究的主要方向之一。Naqvi 等人对原始信号进行去噪后，提取时域标准偏差、偏斜度、峰度等特征和频谱平均值、频谱标准差、梅尔频率倒谱系数等频域特征，将其融合用于后续区分 COPD、肺炎和正常肺音信号<sup>[28]</sup>。Porieva 等人分别利用快速傅里叶变换方法计算呼吸信号的功率谱密度，对肺音信号进行小波变换提取小波系数，用于区分正常肺音、支气管炎和 COPD<sup>[29]</sup>。Zhao 等人提出一种时域和频域结合的肺音信号特征提取方法，首先提取信号的短时平均过零率和短时平均能量，然后提取频谱质心、频谱斜率、频谱对比度、频谱中每帧能量的均方根、色度等频谱特征，再将时域、频率特征融合，充分提取肺音信号的特征进行下一步分类<sup>[30]</sup>。常峥等人利用希尔伯特变换求得肺音信号的边际谱，计算谱能量并求取谱能量的 MFCC，同时提取信号的短时能量，最后将 MFCC 与短时能量融合为新的特征向量<sup>[31]</sup>。Shanthakumari 等人使用连续小波变换得到时频谱图，再从时频谱图中提取能量、熵、方差和波形长度等数学形态特征，用以区分正常肺音和爆裂音<sup>[32]</sup>。Haider 等人提取肺音信号的小波熵和小波包能量特征，最后对健康人、COPD 和哮喘进行分类<sup>[33]</sup>。Rizal 等人对肺音信号分别进行离散小波变换、小波包变换，再计算样本熵、排列熵、近似熵等熵值，然后区分正常支气管、爆裂声、喘息声、胸膜摩擦声和喘鸣音五种肺音信号<sup>[34]</sup>。但是上述方法忽略了肺音信号的时间序列特性，因此本文根据肺音信号的时间序列特性，提取肺音信号的线性预测倒谱系数，根据肺音信号的非线性非平稳特性，提取梅尔频率倒谱系数，再将上述两种特征参数融合，对肺音信号进行更深层的特征提取。

### 1.2.3 肺音信号分类的研究现状

肺音信号分类一直是肺音信号研究的热点，随着研究人员对肺音信号的理论研究日益深入，应用到肺音信号的分类方法也越来越多。其中包括多层感知器、K 最近邻、线性判别器、决策树

和随机森林等机器学习方法。

Lozano 等人根据肺音信号的吸气相，采用集合经验模态分解对肺音信号进行去噪，然后对固有模态分量进行希尔伯特变换得到瞬时频率以及瞬时包络，最后用支持向量机对其进行分类<sup>[35]</sup>。Ulukaya 等人首先对自采数据集进行离散小波变换，然后提取高、低 Q 因子特征子集分别输入 K 最近邻和支持向量机分类器，结果表明高 Q 因子小波分析比低 Q 因子小波具有更高的准确率<sup>[36]</sup>。Fraiwan 等人从肺音信号中提取香农熵、对数能量熵和基于谱图的光谱熵，采用决策树和线性判别分类器作为分类器，使用具有交叉验证损失成本函数的贝叶斯优化进行超参数调整，分类结果优于经典机器学习<sup>[37]</sup>。Lang 等人提取肺音信号的频谱图，再利用半监督支持向量机依据谱图的顶点对正常、噼啪声、喘息声、噪声进行区分<sup>[38]</sup>。Rizal 等人对多个数据集进行幅度归一化，再通过离散小波变换和小波包分解对肺音信号进行分解，利用多层感知器作为分类器，最终结果表明选择所有小波子带比单一特征分类结果更优<sup>[39]</sup>。Sengupta 等人提取肺音信号的倒谱特征，采用带有多层次感知器的人工神经网络分类，鲁棒性更好，并且在低信噪比下也显示出更好的识别精度，且比基于小波变换的方法效果更优<sup>[40]</sup>。但是传统的分类方法只能对少量的特征进行分类，并且分类准确率较低，无法满足实际应用。随着深度学习在生理信号领域的应用，肺音信号的分类准确率也有了很大的提升。

Ma 等人先对肺音信号进行增强，并通过 5 阶巴特沃斯带通滤波器进行去噪，然后提取短时傅里叶变换频谱图，为了提高模型的鲁棒性，在 ResNet 结构中加入非局部层，对肺音信号进行分类<sup>[41]</sup>。Demir 等人提出一种基于卷积神经网络的肺音信号分类方法。先对肺音信号进行短时傅里叶变换，然后使用 VGG16 网络模型提取信号中蕴含的深层特征，对肺音信号进行分类<sup>[42]</sup>。Perna 通过卷积神经网络对肺音信号特征进行分类，为了降低过拟合而采用弹性正则化增强网络的泛化能力，实验结果表明卷积神经网络适用于呼吸系统疾病的分类<sup>[43]</sup>。包善书等人提取肺音信号的对数梅尔频率谱特征，再通过迁移学习对正常肺音和异常肺音进行有效区分<sup>[44]</sup>。Pham 等人将肺音信号的频谱图作为深度学习网络的输入，并对模型性能和模型复杂度进行平衡调节，以达到肺部疾病检测的目的<sup>[45]</sup>。Chen 等人利用深度残差网络对肺音信号的特征学习，并对喘息、爆裂和正常肺音进行分类<sup>[46]</sup>。Nguyen 等人使用协同调整和随机归一化对 ResNet 进行改进，避免了小数据集的过拟合问题，提升肺音信号分类准确率<sup>[47]</sup>。Jayalakshmy 等人首先利用 EMD 提取肺音信号的 IMFs，计算 IMFs 的 Gammatone 倒谱系数，再使用双向长短期记忆神经网络对正常肺音和异常肺音进行区分<sup>[48]</sup>。随着 Transformer 在计算机视觉中取得优异性能之后，其在语音识别中也得到广泛应用。Neto 等人提出一种卷积视觉 Transformer 网络模型对肺音信号进行分类，实验结果表明

模型具有高性能和高准确率<sup>[49]</sup>。Huang 等人提出一种局部循环 Transformer 网络模型，提取肺音信号的特征，对其进行分类<sup>[50]</sup>。Barbouchi 等人以 ResNet50 为主体框架，提取肺部 CT 特征，再经过 Transformer 进行编解码，对肺音进行预测分类<sup>[51]</sup>。Sun 等人将卷积神经网络与 Transformer 结合，提取胸部 CT 图像的特征，获得特征之间远距离的位置关系，更准确地区分肺炎<sup>[52]</sup>。

虽然通过深度学习方法可以实现肺音信号分类，但是网络层数过深会导致模型过拟合。同时肺音信号是时间序列信号，提取其时间序列将提升网络模型的可靠性。因此，本文将卷积神经网络与具有并行结构的 Transformer 结合，提取肺音信号的全局特征对肺音信号进行分类。

### 1.3 本文研究内容

针对慢性阻塞性肺疾病早期筛查困难，导致患者不能及时诊断与治疗等问题，本文采用基于深度学习的方法对肺音信号进行分类研究。论文研究路线如图 1-1 所示。

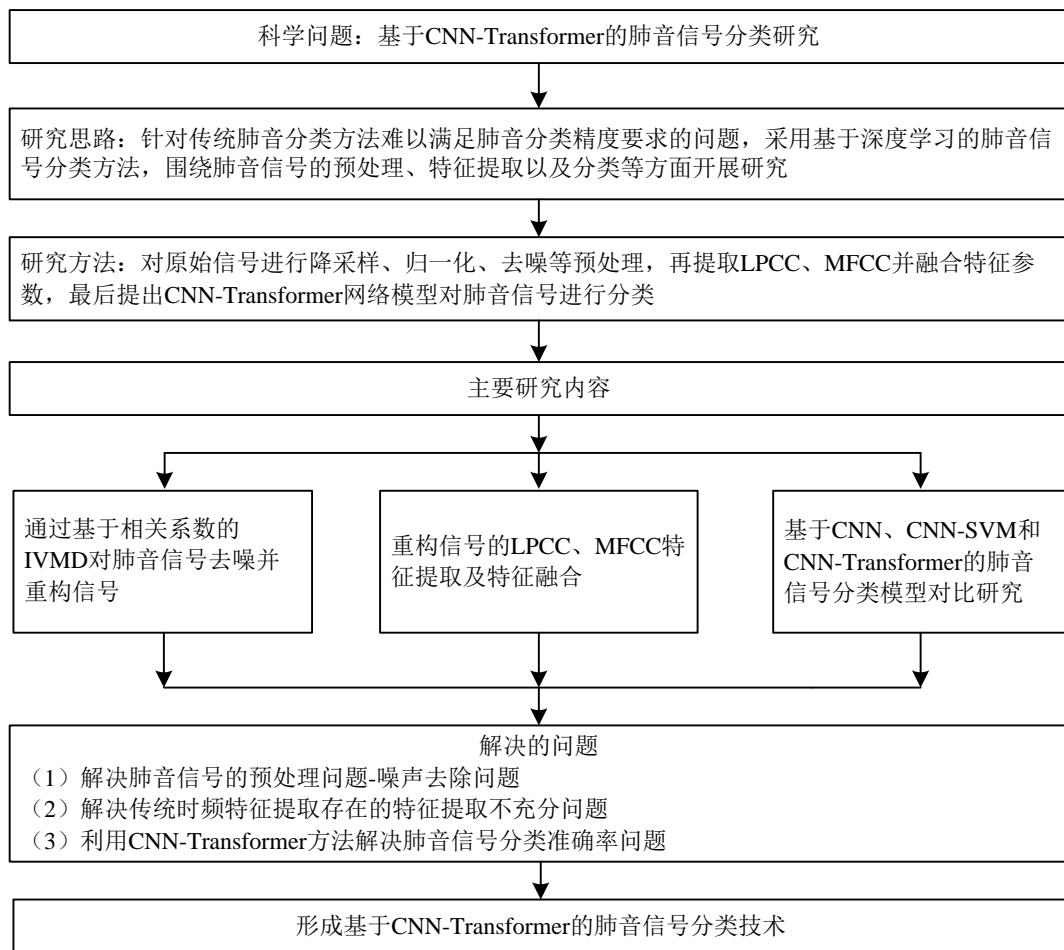


图 1-1 论文研究路线图

首先，为了更高效地处理数据，本文先对原始肺音信号进行降采样、归一化处理，再针对原始信号中包含其他噪声和心音信号的问题，本文采用改进的变分模态分解算法对原始信号进行去

噪处理，使用基于相关系数的方法选取固有模态分量并重构肺音信号，实验表明该方法能够有效去除噪声，提升的肺音信号质量。

其次，针对单一特征不能充分反映肺音信号的问题，为了更全面地提取肺音信号的有用特征，本文根据肺音信号的时间序列特性和非线性非平稳的特性，提取重构肺音信号的线性预测倒谱系数以及梅尔频率倒谱系数，再将两种特征进行融合作为分类特征。

最后，针对目前常用的神经网络模型分类准确率低、网络层数较高的问题，本文利用 Transformer 网络模型具有高并行结构的特点，搭建 CNN-Transformer 网络模型，有效学习时间序列全局特征，更好地将时间序列的前后信息整合到注意机制中，提高分类准确率。

## 1.4 本文结构安排

本文一共分为六个章节，具体章节内容如下：

第一章，主要阐述了基于 CNN-Transformer 的肺音信号分类研究的背景及意义，并介绍了肺音信号预处理的研究现状、肺音信号特征提取的研究现状以及肺音信号分类的研究现状。

第二章，主要介绍了肺音信号分类的理论基础。首先叙述了肺音信号采集、肺音信号类别、肺音信号去噪以及肺音信号特征提取的理论基础，其次对 CNN 以及 Transformer 网络模型用于肺音信号分类的基本理论进行阐述。

第三章，针对在采集过程中不可避免的会混入其他噪声以及心音噪声等问题，采用一种基于改进的变分模态分解的肺音信号去噪算法。首先选择自适应去噪中的 EMD 方法对肺音信号进行去噪，再基于相关系数选取固有模态分量重构信号，针对 EMD 系列存在的模态混叠问题采用 IVMD 对肺音信号进行去噪，最后对实验结果进行对比分析。

第四章，针对目前单一的肺音信号特征提取不能充分保留肺音信号有效成分等问题，提取肺音信号的线性预测倒谱系数和梅尔频率倒谱系数，并将两者融合，得到新的融合特征，提取更丰富的肺音信号特征。

第五章，针对目前常用于肺音信号分类研究方法中迭代速率慢和准确率低等问题，以 CNN 为主体框架，其次利用支持向量机结合 CNN 网络模型以及搭建基于 CNN-Transformer 的肺音信号分类网络模型，最后对分类结果进行对比分析。

第六章，总结基于 CNN-Transformer 的肺音信号分类网络模型的研究工作，并指出本文研究过程中的不足和改进的方法，对今后的研究工作进行展望。

## 第二章 肺音信号预处理及分类的理论基础

肺音信号在采集过程中受到多种因素如心音信号、环境噪声等动态噪声与采集设备的影响，干扰肺音信号的分类结果。因此，去噪成为肺音信号分类的关键环节。本章首先介绍肺音信号的采集以及肺音信号的分类，再阐述肺音信号去噪的理论基础，最后介绍深度学习的相关技术，包括 CNN 和 Transformer 的理论基础。

### 2.1 肺音信号的理论基础

#### 2.1.1 肺音的基本概念及分类

人体在呼吸过程中，气体在呼吸系统内流动引起肺组织摩擦振动，发出声响，这些声响经过肺组织和胸壁传到体表形成肺音。肺音信号是一种非线性非平稳的信号，该信号蕴含着人体丰富的生理、病理信息，可以表征人体呼吸系统的健康状况。根据肺音的听诊部位、声音强度、性质、音调高低以及呼吸气相的长短可分为正常肺音和异常肺音两大类。

正常肺音是人体在健康状况下正常呼吸时产生的声音，根据肺音的采集位置不同分为支气管呼吸音、肺泡呼吸音和支气管肺泡呼吸音三类<sup>[53]</sup>。支气管呼吸音是气体流经支气管形成湍流所产生的声音。肺泡呼吸音是由于肺泡的弹性变化、张力和气体振动使得细支气管和肺泡内的空气流动产生的一种呼吸音。支气管肺泡呼吸音是肺内部的气体在气管中进出产生的声音。当患有呼吸系统疾病时，会导致气管收缩或者阻塞等状况，从而产生异常肺音。

根据声音的持续时间可以将异常肺音分为连续性异常音和间断性异常音两大类<sup>[54]</sup>。连续性异常音是由于气管、支气管部分阻塞导致气体流动时出现湍流所产生的一种持续时间较长的异常肺音。间断性异常音是人体呼吸时的气流使气管或者支气管中的分泌物形成水泡并爆裂产生的一种持续时间较短的异常肺音。

#### 2.1.2 肺音信号的特性

支气管呼吸音在正常人体的锁骨上窝、胸骨上窝、喉部以及颈部均可听到。支气管呼吸音音调较高，响度较大，吸气音比呼气音强，吸气相时长比呼气相短，并且在每个呼吸周期内具有短暂间隔。支气管呼吸音在呼吸气相均可闻及，比肺泡呼吸音沉闷且频率带宽更高。

肺泡呼吸音相对于支气管呼吸音的音调较低，听起来为一种柔和的叹息样的声音，呼吸气相与支气管肺泡呼吸音相比更长。尽管在胸腔大部分区域都存在肺泡呼吸音，但是肺泡呼吸音的频率范围比其他两种正常肺音更小，一般在 100Hz-200Hz<sup>[55]</sup>。肺泡呼吸音的强弱与年龄、性别、体

型、呼吸的深浅、肺组织弹性的大小及胸壁的厚薄等因素有关。因为儿童的胸壁较薄且肺泡弹性比老年人的肺泡弹性强，所以儿童的肺泡呼吸音较老年人强。相对于女性而言，男性呼吸运动的力度较强，且胸壁的皮下脂肪较少，所以男性肺泡呼吸音较女性强。

支气管肺泡呼吸音是空气进出人体气管中发出的声响。吸气时，支气管肺泡呼吸音与正常肺泡呼吸音相比，音调较高；呼气时，支气管肺泡呼吸音与正常肺泡呼吸音相比，其音调较低，声音比较弱并且吸气和呼气之间存在短暂的停顿。支气管肺泡呼吸音存在于吸气相和呼气相，且持续时间接近 1:1。与其他两种正常肺音相比，支气管肺泡呼吸音的能量分布范围广，频率小的为 100Hz，频率大的可达到 5000Hz<sup>[56]</sup>。

哮鸣音和鼾音属于连续性异常音，主要存在于呼气相。这两种连续性异常音与哮喘、支气管炎和 COPD 等疾病有密切的联系，肺音中出现它们通常预示着某种疾病的产生<sup>[57]</sup>。使用电子听诊器和现代信号分析技术可以对其进行分析与客观评估，达到识别不同肺部疾病的目的。它们在时域上的波形类似于正弦波，单个周期持续时间超过 250ms<sup>[58]</sup>，存在这两种异常音的时域波形幅度明显比正常肺音高。它们的频率范围为 100Hz-1000Hz，且频率的细节特征可随着时间的变化越加显著。其中哮鸣音的主频为 400Hz，鼾音的主频为 200Hz<sup>[59]</sup>。哮鸣音的频谱图中所包含的高频成分要明显高于正常肺音。对于非线性非平稳的肺音信号，可以通过频域分析得到更明显的频率相关参数，更有利与后续的信号分析。

啰音又被称为爆裂音，属于间断性异常肺音，是因病变或者呼吸系统阻塞导致的<sup>[60]</sup>。啰音的频率范围为 100~200Hz，持续时长小于 25ms，分为干啰音和湿啰音。干啰音是一种具有音乐性的异常音，音调较高，在呼吸气相均可听到，其强度易改变。湿啰音的音调较低，连续多个出现，在吸气时或吸气末较明显，咳嗽后可出现或消失。啰音出现的位置与疾病具有密切联系，比如出现在两侧肺底，多见于肺淤血、支气管肺炎等疾病；若两肺均出现湿啰音，多为急性肺水肿。

## 2.2 肺音信号预处理的理论基础

### 2.2.1 经验模态分解

美国国家宇航局的华裔科学家黄鄂等人提出经验模态分解<sup>[61]</sup>。该方法作为一种新颖的处理非平稳信号的方法，开启了非线性非平稳信号的自适应时代。EMD 按照不同的时间尺度对信号进行分解，得到一系列的固有模态分量和趋势项。分解为两部分，第一部分是筛选部分，第二部分为重构部分。假设原始信号  $x(t)$  为原始信号，具体分解步骤如下：

步骤一：计算原始信号所有的局部极大值点和极小值点，再通过三次样条插值法拟合所有极

值点，得到信号的上包络线和下包络线，取上包络线和下包络线的均值包络线，如式(2-1)所示。

$$\bar{x}(t) = \frac{x_{up}(t) + x_{down}(t)}{2} \quad (2-1)$$

式中， $x_{up}(t)$  表示上包络线， $x_{down}(t)$  表示下包络线， $\bar{x}(t)$  表示上下包络线的均值包络线。

步骤二：用原始信号  $x(t)$  减去均值包络  $\bar{x}(t)$ ，得到  $y_1(t)$ ，如式(2-2)所示。

$$y_1(t) = x(t) - \bar{x}(t) \quad (2-2)$$

步骤三：筛选出固有模态分量的终止条件有两个：第一个条件是在整个时间序列中，极值点的数量和零点的数量相等或者相差为一；第二个条件是极大值的上包络线和极小值的下包络线在任意时间点上均值为零。若  $y_1(t)$  满足上述两个条件，则记作固有模态分量。如果不满足，则将  $y_1(t)$  作为待处理信号，重复以上步骤，直到  $y_1(t)$  符合固有模态分量条件，此时称  $y_1(t)$  为一个 IMF<sub>1</sub>，记作  $m_1(t)$ ，从原始信号  $x(t)$  中减去  $m_1(t)$  得到  $n_1(t)$ ，如式(2-3)所示。

$$n_1(t) = x(t) - m_1(t) \quad (2-3)$$

步骤四：将所得  $n_1(t)$  作为新的原始信号  $x(t)$ ，重复步骤一、二、三，直到得到所有的 IMF<sub>n</sub>。当  $m_n(t)$  与  $n_n(t)$  满足步骤三的终止条件时分解结束。

经过上述四个步骤，信号由经验模态分解得到一系列的固有模态分量，从 IMF<sub>1</sub> 到 IMF<sub>n</sub> 对应着频率由高到低分布，由此根据不同的固有模态分量组合成不同的频率特性的滤波器。

## 2.2.2 变分模态分解

由于 EMD 系列对肺音信号进行分解时存在模态混叠和端点效应等现象，对肺音信号后续的分析处理产生诸多干扰。为了解决 EMD 方法本身的缺点，加利福尼亚大学的学者 Konstantin Dragomiretskiy 等人对 EMD 进行改进，提出了变分模态分解这一算法<sup>[62]</sup>。该算法以维纳滤波为基础，对信号的变分模型搜寻最优解，根据最优解确定固有模态分量的中心频率和带宽。在一定程度上解决了模态混叠和端点效应等问题。相对于自适应分解算法中的 EMD，变分模态分解算法具有数学理论基础，舍弃递归求解的思路，采用完全非递归求解，分解效率更高，可以自主设定分解固有模态分量的个数。

VMD 算法受启发于 EMD 算法，同样适用于处理非线性非平稳的信号。VMD 算法通过对输入信号进行迭代，将原始信号分解为若干设定个数的固有模态分量。VMD 算法分解过程主要分为两个部分，第一部分是变分问题的构造，第二部分则是对构造的变分问题进行求解。

构建变分问题时运用希尔伯特变换可以得到对应的单边频谱，对肺音信号  $x(t)$  进行希尔伯特变换如式(2-4)所示。

$$H[x(t)] = \frac{1}{\pi} p \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau \quad (2-4)$$

式中， $p$  表示柯西主值，其作用是将信号  $x(t)$  转换为复函数形式，其表达式如式(2-5)所示。

$$A[x(t)] = x(t) + jH[x(t)] = a(t)e^{j\theta(t)} \quad (2-5)$$

式中， $a(t)$  表示瞬时振幅， $\theta(t)$  表示相位，其表达式如式(2-6)、式(2-7)所示。

$$a(t) = \sqrt{x^2(t) + H^2[x(t)]} \quad (2-6)$$

$$\theta(t) = \arctan \frac{H[x(t)]}{x(t)} \quad (2-7)$$

瞬时频率的表达式如式(2-8)所示。

$$f(t) = \frac{1}{2\pi} \cdot \frac{d\theta(t)}{dt} \quad (2-8)$$

将对应的瞬时频率利用时频谱进行表达，其表达式如式(2-9)所示。

$$H(\omega, t) = \operatorname{Re} \sum_{i=1}^n e^{j \int \omega_i(t) dt} \quad (2-9)$$

### (1) 构造变分问题

步骤一：首先对信号进行希尔伯特变换，确定各个固有模态分量的带宽并计算出各个分量  $u_k$  的单边频谱，如表达式(2-10)所示。

$$\left[ (\delta(t) + \frac{j}{\pi t}) \right] u_k(t) \quad (2-10)$$

步骤二：利用表达式(2-10)与负指数函数相乘，将对应的固有模态分量调制到中心频率上，其表达式如式(2-11)所示。

$$\left[ (\delta(t) + \frac{j}{\pi t}) u_k(t) \right] e^{-j\omega_k t} \quad (2-11)$$

步骤三：计算信号的梯度范数，预估固有模态分量的  $u_k(t)$  的带宽，通过计算得到的中心频率和带宽需要约束各个固有模态分量的带宽之和取得最小值，其表达式如式(2-12)所示。

$$\begin{cases} \min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[ (\delta(t) + \frac{j}{\pi t}) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ s.t. \sum_k u_k = f \end{cases} \quad (2-12)$$

式中， $\{u_k\} = \{u_1, u_2, \dots, u_k\}$  中的  $\{u_k\}$  表示经过分解得到的所有固有模态分量的集合， $\{\omega_k\} = \{\omega_1, \omega_2, \dots, \omega_k\}$  表示与之对应  $\{u_k\}$  的中心频率的集合， $j$  表示虚数基本单位， $f$  表示原始肺音信号。

## (2) 变分问题求解

步骤一：为了解决上述约束问题并保证信号绝对可积，需要将上述约束性变分问题转换为非约束性变分问题，引入  $\alpha$  和  $\lambda$  使得计算结果具有更好的收敛性，得到增广拉格朗日函数，其表达式如式(2-13)所示。

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_k \left\| \partial_t \left[ (\delta(t) + \frac{j}{\pi t}) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \quad (2-13)$$

式中， $\alpha$  表示二次惩罚因子，决定着固有模态分量的带宽， $\alpha$  越小，则各个固有模态分量的带宽越大，但是过大的带宽会使得各个分量之间混叠； $\alpha$  越大，则各个固有模态分量的带宽越小，但是带宽过小会使分解过程中部分信号丢失。 $\lambda$  表示拉格朗日算子，在该算法中用于增强上式的约束条件。

步骤二：为了求解非约束变分问题，利用交替方向乘子法不断更新  $u^{n+1}$ 、 $\omega^{n+1}$  与  $\lambda^{n+1}$ ，其中  $u^{n+1}$  的取值更新表达式如式(2-14)所示。

$$u_k^{n+1} = \arg \min_{u_k \in x} \left\{ \alpha \sum_k \left\| \partial_t \left[ (\delta(t) + \frac{j}{\pi t}) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_k u_k(t) + \frac{\lambda(t)}{2} \right\|_2^2 \right\} \quad (2-14)$$

步骤三：对  $u_k^{n+1}$  进行傅里叶等距变换将其从时域转换为频域形式，变换后表达式如式(2-15)所示。

$$\hat{u}_k^{n+1} = \arg \min_{\hat{u}_k, u_k \in x} \left\{ \alpha \left\| j(\omega - \omega_k) \left[ (1 + \text{sgn}(\omega + \omega_k)) \hat{u}_k(\omega) \right] \right\|_2^2 + \left\| f(\omega) - \sum_i \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2} \right\|_2^2 \right\} \quad (2-15)$$

步骤四：若用  $\omega - \omega_k$  表示上式中的  $\omega$ ，则根据信号对称性可得到非负频率区间的积分形式，其表达式如式(2-16)所示。

$$\hat{u}_k^{n+1} = \arg \min_{\hat{u}_k, u_k \in x} \left\{ \int_0^\infty 4\alpha(\omega - \omega_k)^2 \left| \hat{u}_k(\omega) \right|^2 + \left| \hat{f}(\omega) - \sum_i \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2} \right|^2 d\omega \right\} \quad (2-16)$$

此时最优解如式(2-17)所示。

$$\hat{u}_k(\omega) = \frac{\hat{f}(\omega) - \sum_k \hat{u}_k(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \quad (2-17)$$

同理可得  $\omega_k^{n+1}$  与  $\lambda^{n+1}$  的最优解表达式如式(2-18)、式(2-19)所示。

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left| \hat{u}_k(\omega) \right|^2 d\omega}{\int_0^\infty \left| \hat{u}_k(\omega) \right|^2 d\omega} \quad (2-18)$$

$$\lambda^{n+1}(\omega) = \lambda^n(\omega) + \tau \left[ \hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega) \right] \quad (2-19)$$

式中， $\tau$  表示一个拉格朗日乘子的更新因子， $n$  表示迭代次数，且  $n = \{1, 2, \dots, N\}$ 。

步骤五：不断重复上述步骤，直到符合终止条件时停止迭代，得到  $k$  个 IMF 分量，终止迭代条件表达式如式(2-20)所示。

$$\frac{\sum_k \left\| \hat{u}_k^{n+1} - \hat{u}_k^n \right\|_2^2}{\left\| \hat{u}_k^n \right\|_2^2} < e \quad (2-20)$$

VMD 算法具体过程如下：

步骤一：初始化  $\{u_k^1\}$ ,  $\{\omega_k^1\}$ ,  $\lambda^1$  和  $n=0$ ;

步骤二：令  $n=n+1$ , 进入循环;

步骤三：根据  $u_k^{n+1}$  和  $\omega_k^{n+1}$  的更新公式更新，直到分解的固有模态分量个数达到 K 时，停止内循环；

步骤四：根据  $\lambda^{n+1}$  的更新公式进行更新；

步骤五：循环步骤二至步骤四，直到满足式(2-26)所示的收敛条件。

## 2.3 基于深度学习的肺音信号分类模型理论基础

随着计算机等工具的发展，人工智能领域得到飞速进步。深度学习模拟人脑思维方式，对大量的数据进行自动学习并获取高层特征，提升网络模型的鲁棒性。对比人工主观获得特征的浅层网络模型，深度学习拥有多层网络模型，可以更客观地获得高层特征，减少主观因素的影响。深度学习的基本结构包含输入层、隐含层、输出层等，深度学习的工作期间，将数据通过输入层输入，经过隐含层，直到输出层再逐层反向传播至输入层，通过循环迭代，可以减少训练过程中的特征损失，提高准确率，最后输出学习的信号特征。

### 2.3.1 卷积神经网络的理论基础

卷积神经网络成为最流行的深度学习架构之一，推动着肺音信号分类的发展。卷积神经网络的基本结构由输入层、卷积层、池化层、激活函数层、全连接层、Softmax 层和输出层几个部分组成。卷积神经网络模型图如图 2-1 所示。

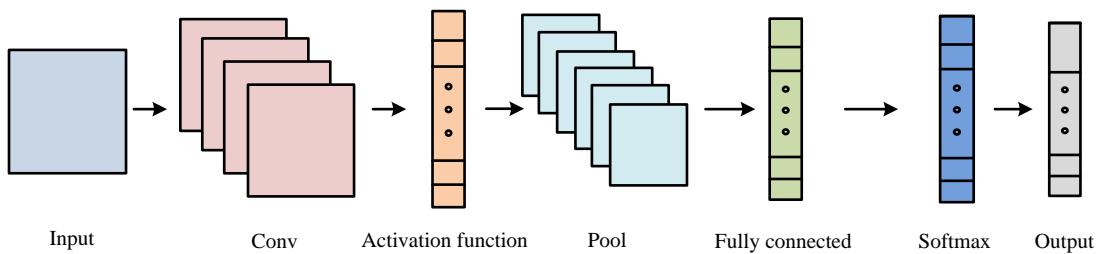


图 2-1 卷积神经网络模型图

在肺音信号分类任务中，将预处理后的信号由输入层依次经过卷积神经网络中的卷积层、激活函数和池化层提取肺音信号的特征，再经过全连接层和 Softmax 层对肺音信号进行分类，这是前向传播的过程。通过逐层向前的反向传播，不断更新迭代权重参数达到降低判断误差的目的。

输入层是将输入信号输送到网络模型中进行特征提取，直到输出层输出分类结果。在输入层中，输入信号进行预处理，将数据长度、采样率进行统一操作，缩短训练时间，设置合理的激活函数值域，保证网络的正常运行。

卷积层是 CNN 的主要结构元素之一，其目的是对输入数据进行特征提取。卷积层中有多个可以学习的卷积核，前一层输出的特征经过卷积核进行卷积运算，即输入信号与卷积核进行点积运算，再经过激活函数激活，得到特征。卷积核相当于大脑的神经元，每一个神经元都连接着前一层中距离近的多个神经元。深度卷积层中，输入层与卷积核的卷积结果会生成更抽象更高层的特征映射，从而识别出肺音信号分类任务中最具辨别的特征。卷积运算的理论数学表达式如式(2-21)所示。

$$y(n) = (f * u)(n) = \int_{-\infty}^{+\infty} f(x)u(n-x)dx \quad (2-21)$$

式中,  $f(x)$  表示输入信号,  $u$  表示特征提取的核函数,  $y(n)$  表示输出。

因为肺音信号是数字信号, 时间离散化, 因此卷积运算的实际表达式如式(2-22)所示。

$$y(n) = (f * u)(n) = \sum_{x=-\infty}^{+\infty} f(x)u(n-x) \quad (2-22)$$

激活函数通常在卷积层后面, 其作用是对上一层的特征进行非线性映射传递给下一层, 增强卷积神经网络的拟合能力。激活函数具有非线性、单调性、连续可微性, 可以分为饱和激活函数和非饱和激活函数。在无穷处, 若激活函数的导数趋于 0 则称为饱和激活函数, 典型的饱和激活函数包括 Sigmoid 函数、Tanh 函数, 非饱和激活函数有 ReLU 函数等。其中 Sigmoid 函数的输出映射在 [0,1] 之间, Tanh 函数的输出映射在 [-1,1] 之间, Tanh 比 Sigmoid 收敛速度快, 优化稳定, 但计算量大, 容易产生梯度消失。ReLU 函数在梯度下降时, 收敛速度比 Sigmoid 函数和 Tanh 函数更快, 只需要一个阈值, 计算速度更快, 有效解决 Sigmoid 函数梯度消失的问题。

池化层通过对卷积后提取的特征进行非线性下采样, 保留明显的信号特征, 从而可以压缩数据, 减少参数量, 降低网络的计算复杂度和成本避免出现过拟合。池化过程是用与卷积核大小相同的区域块, 按照设置的步长由左向右、由上到下依次滑动, 常用的池化操作有平均池化和最大池化。平均池化是对滑动窗口区域内的元素取平均值, 将平均值作为下一层的输入。最大池化是对移动窗口区域内的元素取最大值。平均池化和最大池化的过程如图 2-2 所示, 其中滑动窗口设置为  $2 \times 2$ , 步长设置为 2。

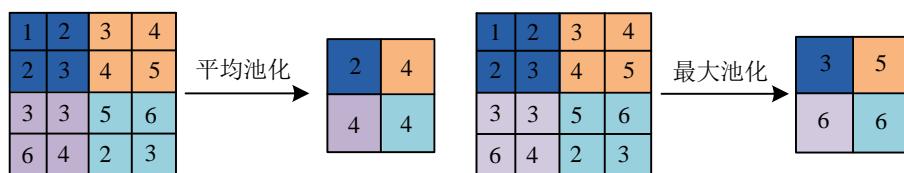


图 2-2 平均池化和最大池化的过程

信号从输入层依次经过卷积、激活、池化操作后得到高度抽象的特征, 而全连接层可以将输入的二维特征转换成一维特征向量, 对特征进行整合并归一化处理。全连接层相当于一个简单的前馈神经网络, 一般位于网络的末端, 全连接层中每一层的结点都与上一层的所有结点相互连接。

Softmax 层又称为分类层, 位于全连接层之后, 根据全连接层输出值, 对每种分类情况输出一个概率。概率计算如式(2-23)所示。

$$P_k = \frac{\exp(c_k)}{\sum_{i=1}^n \exp(c_i)} \quad (2-23)$$

式中,  $n$  表示输出层中神经元的个数,  $P_k$  表示第  $k$  个神经元的输出,  $c_i$  表示输入信号。

输出层用来显示最终的分类预测结果, 在不同的任务中, 输出的结果也各不相同。例如在分

类任务中，最终输出的是具体的分类结果。在输出层不仅可以引入 dropout 操作，随机删除神经网络中的部分神经元，还可以引入局部归一化、数据增强等操作增加网络模型的鲁棒性。

### 2.3.2 Transformer 网络模型的理论基础

Transformer 网络模型是由 Google 团队于 2017 年提出的<sup>[63]</sup>，该模型采用 Encoder 与 Decoder 的架构，极大地缩短了神经网络的训练时间。Transformer 网络模型图如图 2-3 所示。

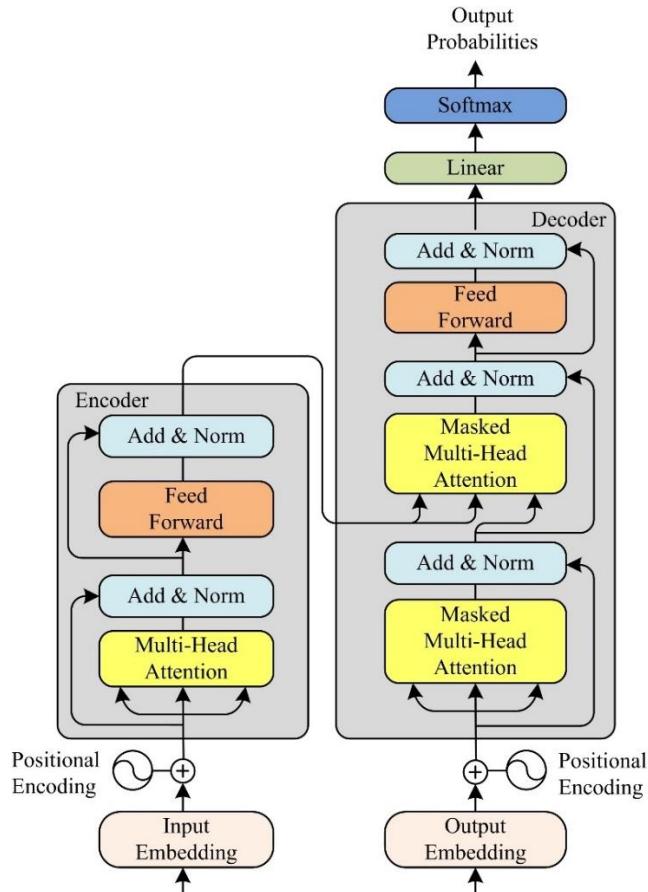


图 2-3 Transformer 网络模型图

CNN 在卷积层中只有与卷积核大小相匹配的信息才可以相互交互，对于长时间序列则需要具有更深层的网络提取特征，而层数过多会导致网络过拟合，影响后续分类结果。RNN 每次运算按照输入时间序列的顺序进行处理，在 T 时刻的计算依赖 T-1 时刻的计算结果，处理方式不灵活，导致训练以及推理的速度较慢。Transformer 的提出解决了上述两种网络模型在时间序列上的缺点，利用左侧的 Encoder block 和右侧的 Decoder block 并行处理输入信号，大大缩减了训练时间。

图 2-3 中输入 Input Embedding 后需要给每个时间序列添加位置编码(Positional Encoding)，这样做是因为时间序列中时间顺序非常重要，同一个时间序列点在不同的位置出现，含义可能发生翻天覆地的变化。可见获取时间序列点出现在整段时间序列中的位置信息是一件很重要的事情。

但是 Transformer 模型是完全基于自注意力机制，而自注意力机制不能获取时间序列点的位置信息，即便是打乱整段时间序列中时间序列点的位置，每个时间序列点还是可以与其他时间序列点之间计算注意力值。它相当于一个功能强大的时间序列模型，对结果没有任何影响。所以在输入的时候需要给每一个时间序列点添加位置编码。

Transformer 网络模型依赖自注意力机制，Multi-Head Attention 由多个自注意力机制(Self-Attention)组成，自注意力机制令源序列和目标序列相互关联，使得源序列和目标序列的 Embedding 蕴含更丰富的信息。Self-Attention 和 Multi-Head Attention 结构图如图 2-4 所示。

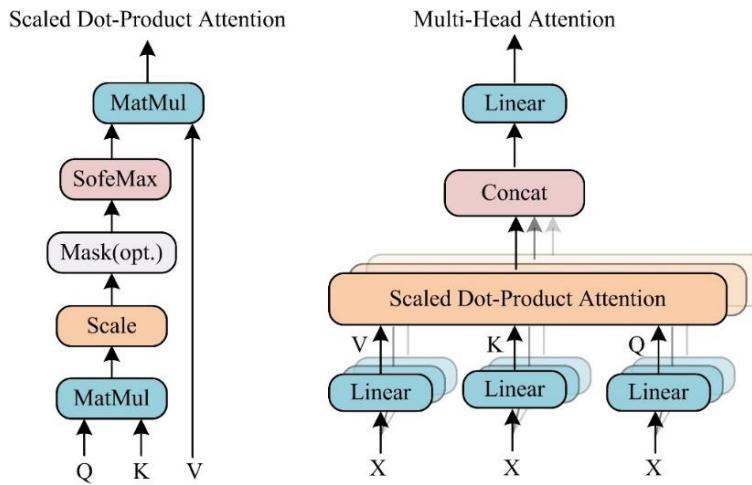


图 2-4 Self-Attention 和 Multi-Head Attention 结构图

Self-Attention 的具体计算过程如式(2-24)所示。

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2-24)$$

式中，Q、K、V 分别是 Query、Key、Value 这三个矩阵的缩写。Query 和 Key 作用之后得到 Attention 的权值，该权值再与 Value 作用得到 Attention 的值。 $d_k$  表示矩阵 Q、K 的列数，即向量维度。Multi-Head Attention 是将每一个 Self-Attention 的输出拼接在一起，然后传给 Linear 层，进行线性变换得到最终的输出。

Add & Norm 层是由 Add 和 Norm 两部分组成的。其中 Add 表示残差连接，主要作用是解决多层网络训练的问题，防止在深度神经网络训练时出现退化。退化是指随着神经网络的层数增加，Loss 逐渐减小，在达到某一值后趋于饱和，继续增加网络层数反而导致 Loss 增大。Norm 表示 Layer Normalization(LN)，主要用于归一化每一层的激活值，加快网络收敛速度，提升网络的稳定性。不选用 Batch Normalization(BN)是因为 BN 是在同一个 batch 中的不同样本之间的同一位置神经元进行的归一化，是对相同的维度进行归一化，特别依赖 batch size；而 LN 是在同一样本中不

同神经元之间进行的归一化，即对同一层神经网络的输出做一个归一化，不依赖 batch size。

Feed Forward 层是一个两层的全连接层，先线性变换然后进行 Relu 激活，再进行线性变换。通过该层可以将输入信号映射到更高维度的空间中，再经过 Relu 激活函数筛选特征，筛选结束变回原来的维度，再次经过 Add & Norm 层，输入下一个 Encoder block 中，通过所有 Encoder block 后再输入 Decoder block 中。

Decoder block 中的 Masked Multi-Head Attention 计算原理与 Encoder block 中的 Multi-Head Attention 一样，其中 mask 码表示掩码，可以掩盖某些值，使其在参数更新时不产生任何效果。Transformer 中有两种掩码，分别是 Padding Mask 和 Sequence Mask。其中 Padding Mask 是为了将输入序列进行对齐，对于较短的序列进行补零操作，若是输入序列太长，则只截取左边的内容，舍弃多余的序列；Sequence Mask 的作用是令 Decoder block 不能获取未来的信息，即在 T 时刻的解码输出只能依据 T 时刻之前的输出，而不能依据 T 时刻之后的输出，因此通过 Sequence Mask 达到隐藏 T 时刻后面的信息的效果。最后经过线性变换再通过 Softmax 层得到输出的概率分布。

## 2.4 本章小结

本章主要对肺音信号的预处理和分类的理论基础进行了阐述。首先介绍了肺音信号的理论基础，包括肺音的基本概念和肺音信号的特性。其次，介绍了肺音信号预处理的理论基础，包括经验模态分解和变分模态分解。最后介绍了深度学习中的卷积神经网络和 Transformer 的基本网络模型，并对其基本结构进行了详细的说明，为后续章节奠定理论基础。

## 第三章 肺音信号的采集及预处理

本章对肺音信号的采集以及预处理等工作进行详细的论述。首先对肺音信号进行降采样和归一化处理，然后通过经验模态分解对肺音信号进行去噪，采用基于相关系数的固有模态分量选取方式。针对经验模态分解不彻底的问题，采用改进的自适应噪声集合经验模态分解算法。最后针对 EMD 系列的端点效应等问题，采用鲁棒性更佳、理论支撑更完备的变分模态分解，并对比了不同种类肺音信号预处理后的差异。

### 3.1 肺音信号的采集

肺音信号采集系统由硬件和软件两部分组成。硬件部分主要由：电子听诊器、外置声卡以及嵌入式开发板构成。软件部分由操作环境和应用软件构成。肺音信号采集是重要的环节，选择具有采集和放大功能的 E-ScopeII 电子听诊器作为采集设备，可以减少部分噪声，Creative Surround5.1 作为外置声卡对采集的肺音信号进行 A/D 转换，采用 Jetson AGX Xavier 开发板作为上位机，对后续数据分析提供支撑。在采集过程中，采样率统一设置为 44.1kHz，健康人先在室温静坐 3 分钟，待信号稳定后采集 10 个周期，健康人群约 2-4 秒呼吸一次，因此，肺音采集的时间约为 30 秒，将采集的肺音信号传输至上位机，以.wav 格式进行存储。

电子听诊器采用的是 E-Scope II 电子听诊器，如图 3-1(a)所示。该电子听诊器包括拾音器、放大和滤波电路、处理芯片、音量调节、切换开关等部分。其中，拾音器的主要作用是采集肺音；放大及滤波电路的作用是将肺音信号进行初步过滤和放大；处理芯片用于降低杂音的干扰，获取杂声较少的肺音信号；音量调节的作用是增大或者减小听诊声音；切换开关可以转换电子听诊器的听诊频率。该电子听诊器不仅可以放大肺部的声音，而且可以通过滤波电路滤除部分环境噪声，在提高肺音信号采集质量的同时，克服了传统听诊器噪音高的缺点。E-Scope II 电子听诊器具有第二信号输出接口，可以将采集到的肺音信号上传至上位机中进行存储，它对肺音信号的采集质量具有重要的作用。

声卡采集器的作用是消除内录干扰，保证储存在上位机中的信号质量，如图 3-1(b)所示。由于上位机内置的声卡在录制过程中容易受到机器内部噪声干扰，所以最终选择使用外置声卡 Creative Surround 5.1 进行 A/D 转换，以此来降低机器干扰，提高肺音信号的质量。

上位机主要是为了后续的数据分析并应用到实际临床中，如图 3-1(c)所示。本文使用英伟达公司的 Jetson AGX Xavier 开发板作为上位机，配置 Ubuntu 系统。该上位机具有电脑所有的基本

功能，支持运行深度学习网络模型，可实现图像分类和目标检测等功能，相对于常用电脑而言，是一种低功耗、边缘计算的设备。该上位机搭载了 8 核 ARM v8.2 64 位 CPU，具有 32GB 的内存，可以满足实验的硬件需求。该主板四周具有 2 个 USB3.1 接口和 1 个 HDMI 接口，可以用来连接外部输入设备和外接显示屏便于操作和展示实验结果。音频传输部分有 3.5mm 的音频接口，可以接入麦克风、耳机和扬声器等设备，方便对肺音信号进行输入和输出。此款上位机与前几代产品相比，外形更加小巧，节能高效，方便携带，支持更多的功能。该上位机最大驱动电流为 2.5A，能够满足外置声卡和外置显示器等设备的供电需求；系统更加稳定，适用于便携的医疗设备。



图 3-1 硬件部分构成

Audacity 软件是一款常用于音频处理的软件，功能较为全面，输入与输出均支持.wav 声音格式，并且采样率可以根据用户需求进行设置。在本文中主要用于录制并保存肺音信号，肺音信号采集系统图如图 3-2 所示。

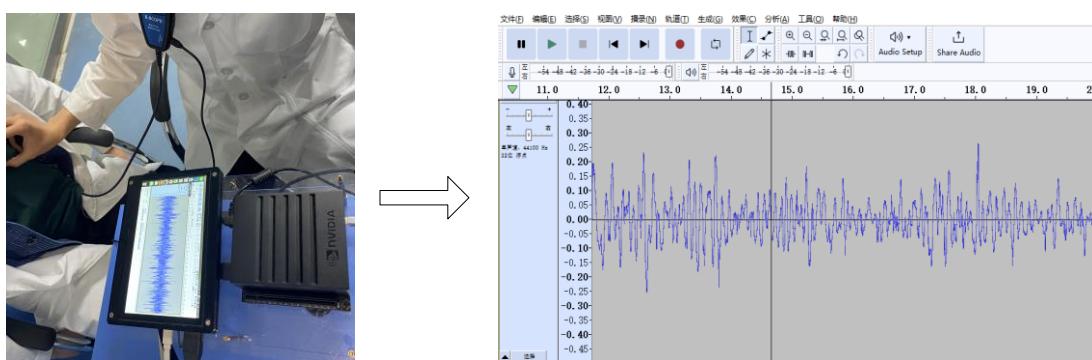


图 3-2 肺音信号采集系统图

### 3.2 肺音信号的预处理

由 EMD 分解的两个限定条件可知，其适用于采样点较少的肺音信号，然而肺音信号采样率一般较大，不能直接使用 EMD 分解，所以需要对原始肺音信号进行降采样以及归一化处理。原始肺音信号采样率为 44100Hz，使用 MATLAB 的 resample 函数对原始肺音信号进行重采样，重采样率为 4410Hz。由于肺音信号的单个周期为 2s 左右，所以选择降采样后的 12000 个采样点

作为后续处理的数据。

肺音信号降采样前后波形对比图如图 3-3 所示，其中，图 3-3(a)是健康人降采样前后波形对比图，图 3-3(b)是 COPD 患者降采样前后波形对比图，从波形可以直观地得出降采样不会改变原始信号的基本特征。降采样前后肺音信号功率对比如表 3-1 所示，后续可根据信号的功率对去噪效果进行评价。

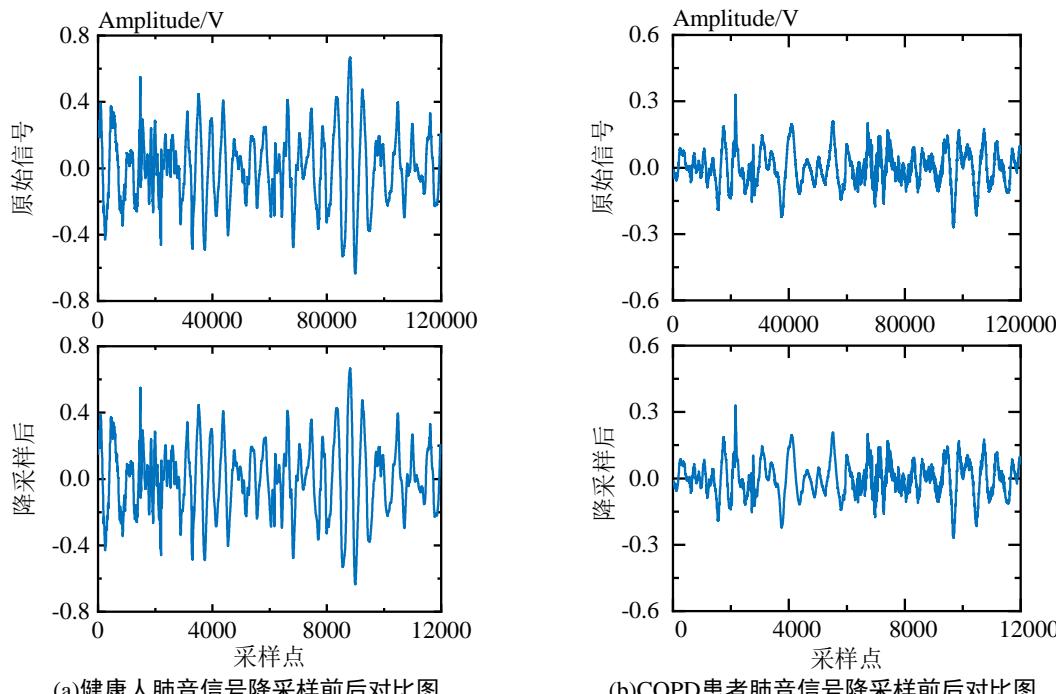


图 3-3 肺音信号降采样前后波形对比图

表 3-1 降采样前后肺音信号功率对比表

样本	原始信号功率	降采样后功率
健康人样本一	0.0471	0.0471
健康人样本二	0.0007	0.0007
健康人样本三	0.0048	0.0047
COPD 患者一	0.0065	0.0065
COPD 患者二	0.0092	0.0092
COPD 患者三	0.0467	0.0467

从图 3-3 中可以看出，健康人肺音信号的幅度大于 COPD 患者的肺音信号幅度，可能是因为健康人呼吸平稳有力，清晰地传递到听诊器；而 COPD 患者呼吸断断续续，信号微弱，传递到听诊器的信号有限。同时可以得出原始的肺音信号波形图中存在毛刺，这是因为在采集过程中混有心音和环境噪声，影响肺音信号本身的特征，所以不能直接对肺音进行特征提取。因此，在进行

特征提取前需要对原始肺音信号进行去噪处理。EMD 在噪声的处理上具有自适应的特点，可以去除环境噪声，并利用心音和肺音主频差异去除心音，从而取得较好的去噪效果。从表 3-1 中可以得出降采样前后肺音信号功率变化可忽略不计，表明降采样前后并不改变肺音信号的特征。

### 3.3 基于 EMD 的肺音信号去噪研究

根据第二章 EMD 的理论基础部分可得经验模态分解的算法流程图如图 3-4 所示。

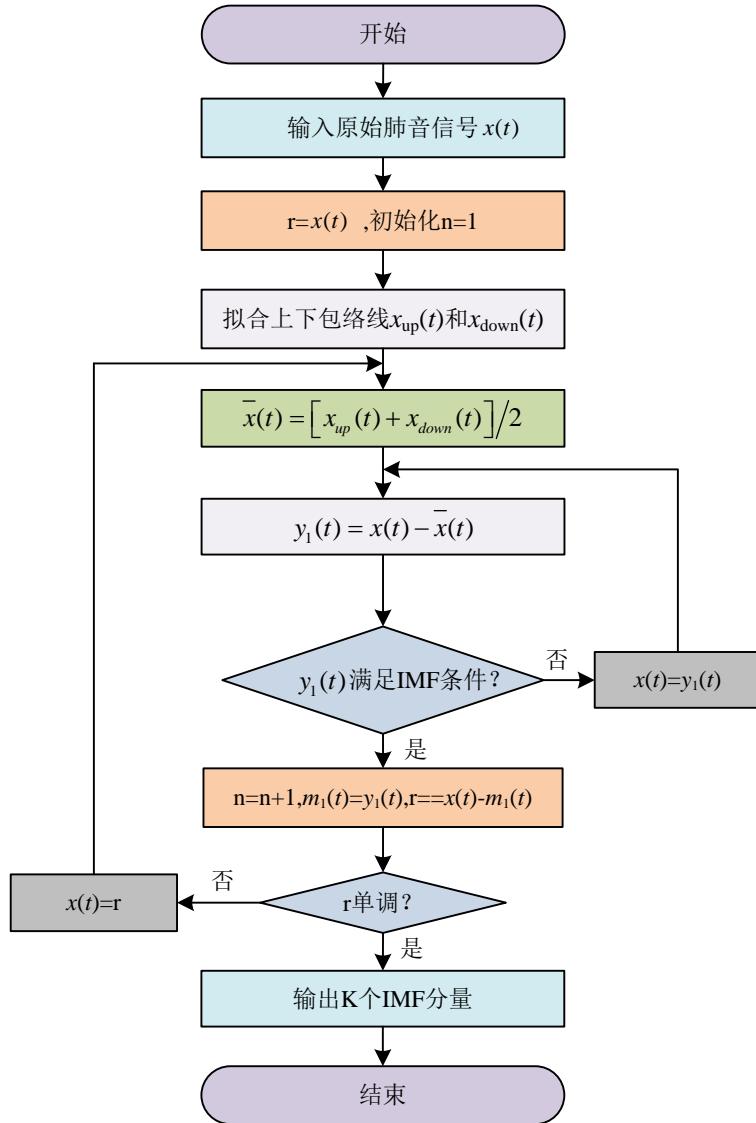


图 3-4 经验模态分解的算法流程图

首先对降采样后的肺音信号进行 EMD 分解，EMD 分解图如所示。其中，图 3-5(a)是健康人肺音信号的 EMD 分解图，图 3-5(b)是 COPD 患者肺音信号的 EMD 分解图。由图 3-5 可知，EMD 分解得到的固有模态分量按照频率由高到低排列。图 3-5(a)中  $IMF_1$ 、 $IMF_2$  表示高频噪声， $IMF_6$ 、 $IMF_7$  表示低频噪声， $IMF_8$  表示肺音信号的趋势；图 3-5(b)中  $IMF_1$ 、 $IMF_2$  为高频噪声， $IMF_7$  表示

低频噪声,  $\text{IMF}_8$  是信号的残差分量, 表示肺音信号的趋势。求取各个固有模态分量与原始信号的相关系数如表 3-2 所示。

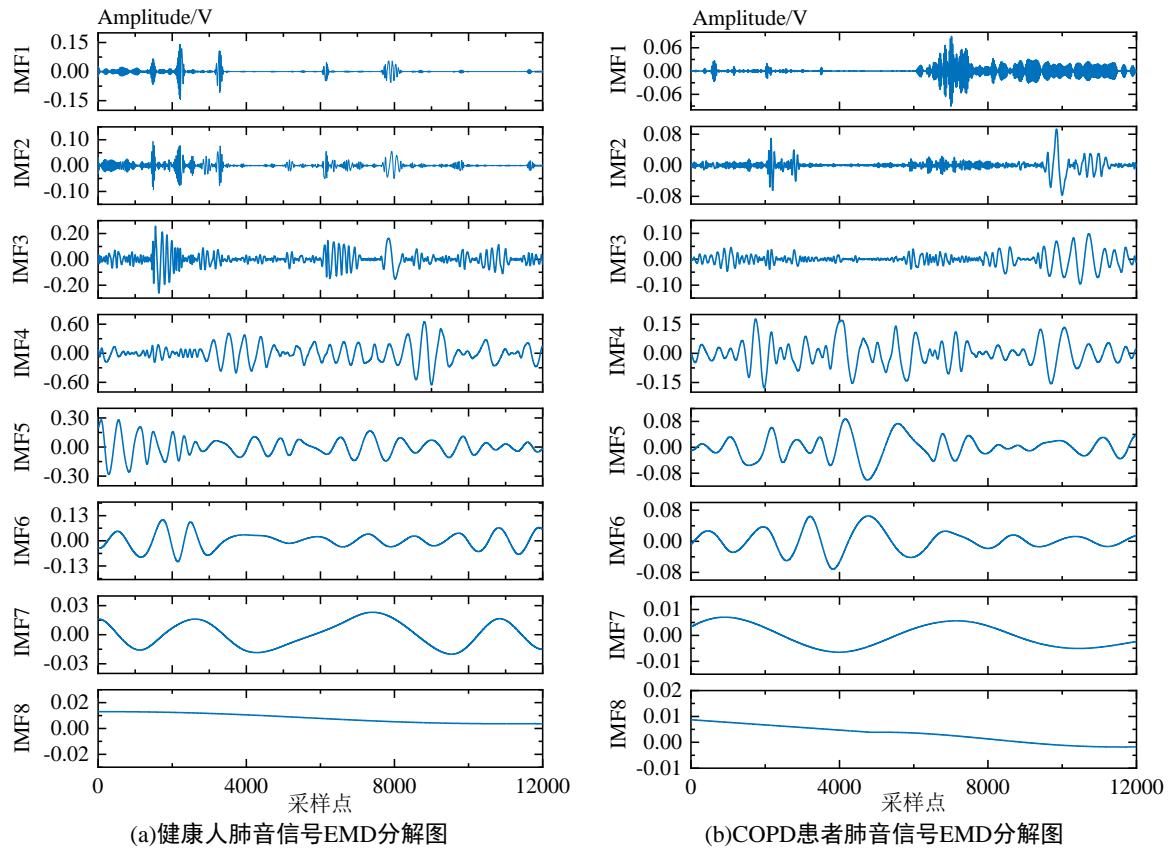


图 3-5 肺音信号 EMD 分解图

表 3-2 EMD 分解各个固有模态分量相关系数

固有模态分量条数	健康人样本一	COPD 患者一
IMF <sub>1</sub>	0.0401	0.0578
IMF <sub>2</sub>	0.0765	0.1947
IMF <sub>3</sub>	0.2410	<b>0.4192</b>
IMF <sub>4</sub>	<b>0.8593</b>	<b>0.7730</b>
IMF <sub>5</sub>	<b>0.4398</b>	<b>0.3198</b>
IMF <sub>6</sub>	0.0902	0.1393
IMF <sub>7</sub>	0.0495	0.0137
IMF <sub>8</sub>	0.0037	0.0088

由表 3-2 中的结果可知, 健康人肺音信号中的  $\text{IMF}_1\text{-}\text{IMF}_3$ 、 $\text{IMF}_6\text{-}\text{IMF}_8$  和 COPD 患者中的  $\text{IMF}_1$ 、 $\text{IMF}_2$ 、 $\text{IMF}_6\text{-}\text{IMF}_8$  的相关系数较小, 选取相关系数大于等于 0.3 的固有模态分量进行信号重构。因此, 选取健康人肺音信号的  $\text{IMF}_4$ 、 $\text{IMF}_5$  重构信号, 重构肺音信号波形图如图 3-6(a)所示。选取 COPD 患者肺音信号的  $\text{IMF}_3$ 、 $\text{IMF}_4$ 、 $\text{IMF}_5$  重构信号, 重构肺音信号波形图如图 3-6(b)所示。与原

始信号图相比，去除噪声后，信号波形明显平滑，毛刺减少。

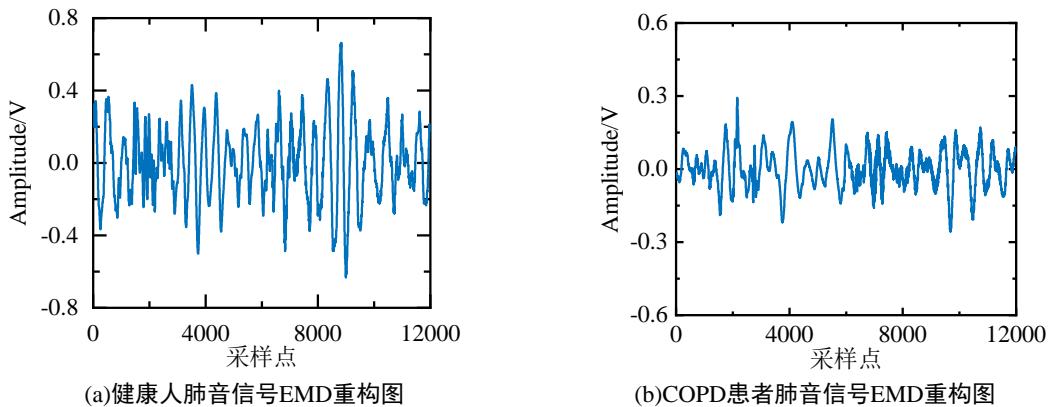


图 3-6 EMD 分解后重构肺音信号波形图

### 3.4 基于 IVMD 的肺音信号去噪研究

改进的变分模态分解算法(IVMD)算法流程图如图 3-7 所示。

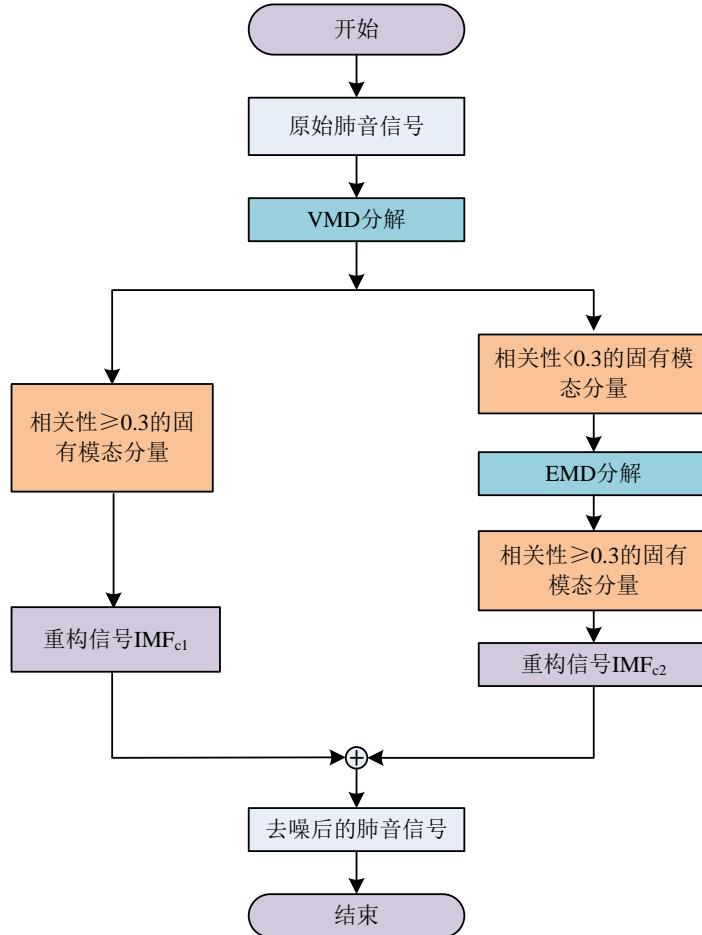


图 3-7 IVMD 分解算法流程图

步骤一：设定 K 值为 6，确定 VMD 算法分解时的固有模态分量的分解个数。

步骤二：从分解出的固有模态分量中选取相关系数大于等于 0.3 的分量重构为  $IMF_{c1}$ ，剩余

的固有模态分量重构为待分解的信号  $x'(t)$ 。

步骤三：将步骤二中待分解的信号  $x'(t)$  进行 EMD 分解，分解之后选取相关系数大于等于 0.3 的固有模态分量重构为  $\text{IMF}_{c2}$ ，将其余固有模态分量舍弃。

步骤四：将步骤二中的  $\text{IMF}_{c1}$  与步骤三中的  $\text{IMF}_{c2}$  叠加，得到最终去除噪声的肺音信号。

对 VMD 分解后的固有模态分量进行 EMD 分解，进一步去除了噪声成分，保留了信号中的有效成分，从整体上大大减少了噪声对肺音信号的干扰，提升了肺音信号的质量，方便后续更精确地提取肺音信号的特征。具体过程如下：首先对降采样后的肺音信号进行 VMD 分解，如图 3-8 所示，图 3-8(a)是健康人样本的 VMD 分解图，图 3-8(b)是 COPD 患者样本的 VMD 分解图。

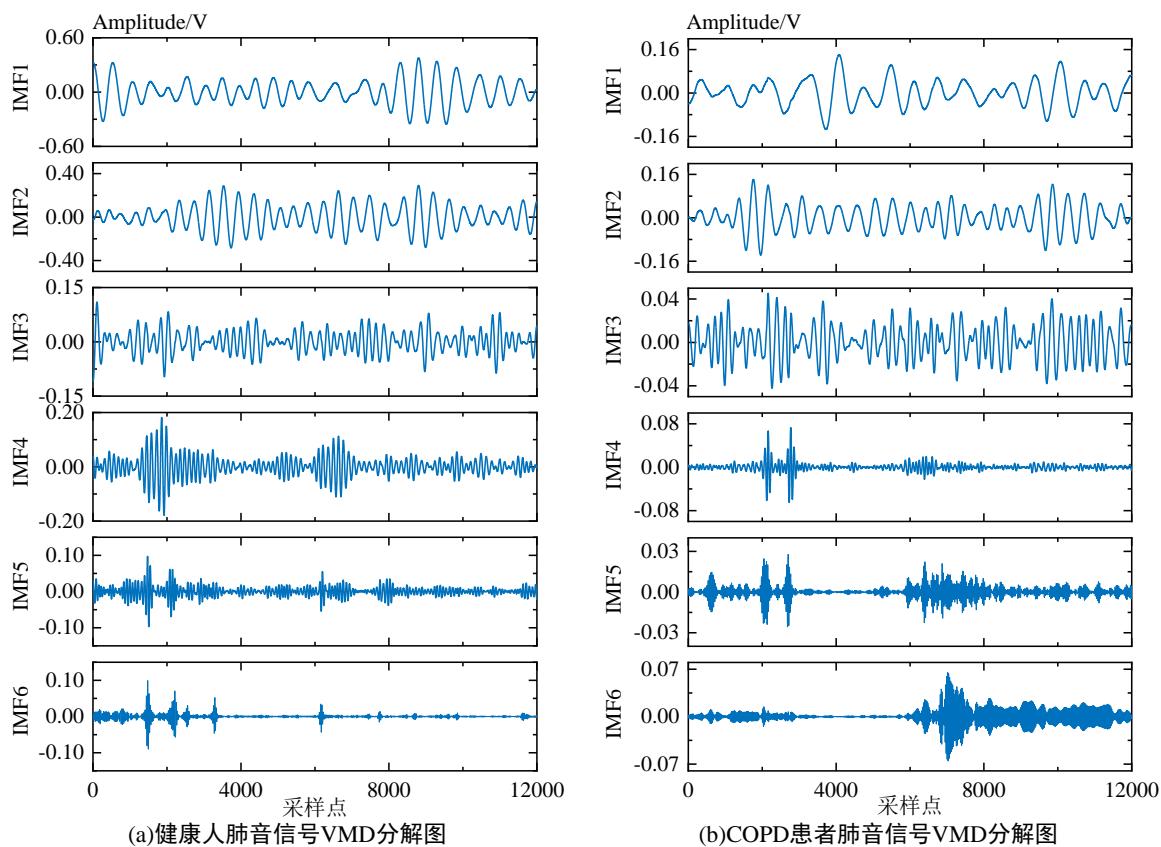


图 3-8 肺音信号 VMD 分解图

由图 3-8 可知，肺音信号经 VMD 分解后，得到的固有模态分量按照频率由低到高排列，(a) 和(b)中  $\text{IMF}_5$ 、 $\text{IMF}_6$  表示高频噪声。VMD 分解的各个固有模态分量与原始信号的相关系数如表 3-3 所示。由表 3-3 中的结果可知，健康人和 COPD 患者固有模态分量中的  $\text{IMF}_5$ 、 $\text{IMF}_6$  相关系数较小，均为噪声，选取相关系数大于等于 0.3 的固有模态分量进行信号重构。因此，选取健康人的  $\text{IMF}_1$ 、 $\text{IMF}_2$  重构信号为  $\text{IMF}_{c1}$ ，将其余的  $\text{IMF}_3$ 、 $\text{IMF}_4$ 、 $\text{IMF}_5$ 、 $\text{IMF}_6$  合并为待分解信号  $x_1(t)$ ；选取 COPD 患者的  $\text{IMF}_1$ 、 $\text{IMF}_2$ 、 $\text{IMF}_3$  重构为  $\text{IMF}'_{c1}$ ，将剩余的  $\text{IMF}_4$ 、 $\text{IMF}_5$ 、 $\text{IMF}_6$  合并为待分解

信号  $x_1'(t)$ , 并对  $x_1(t)$  和  $x_1'(t)$  进行 EMD 分解, 得到分解图如图 3-9 所示。

表 3-3 VMD 分解各个固有模态分量相关系数

固有模态分量条数	健康人样本一	COPD 患者一
IMF <sub>1</sub>	<b>0.8047</b>	<b>0.7226</b>
IMF <sub>2</sub>	<b>0.7306</b>	<b>0.7397</b>
IMF <sub>3</sub>	0.2550	<b>0.3524</b>
IMF <sub>4</sub>	0.2201	0.1628
IMF <sub>5</sub>	0.1145	0.1041
IMF <sub>6</sub>	0.0063	0.1563

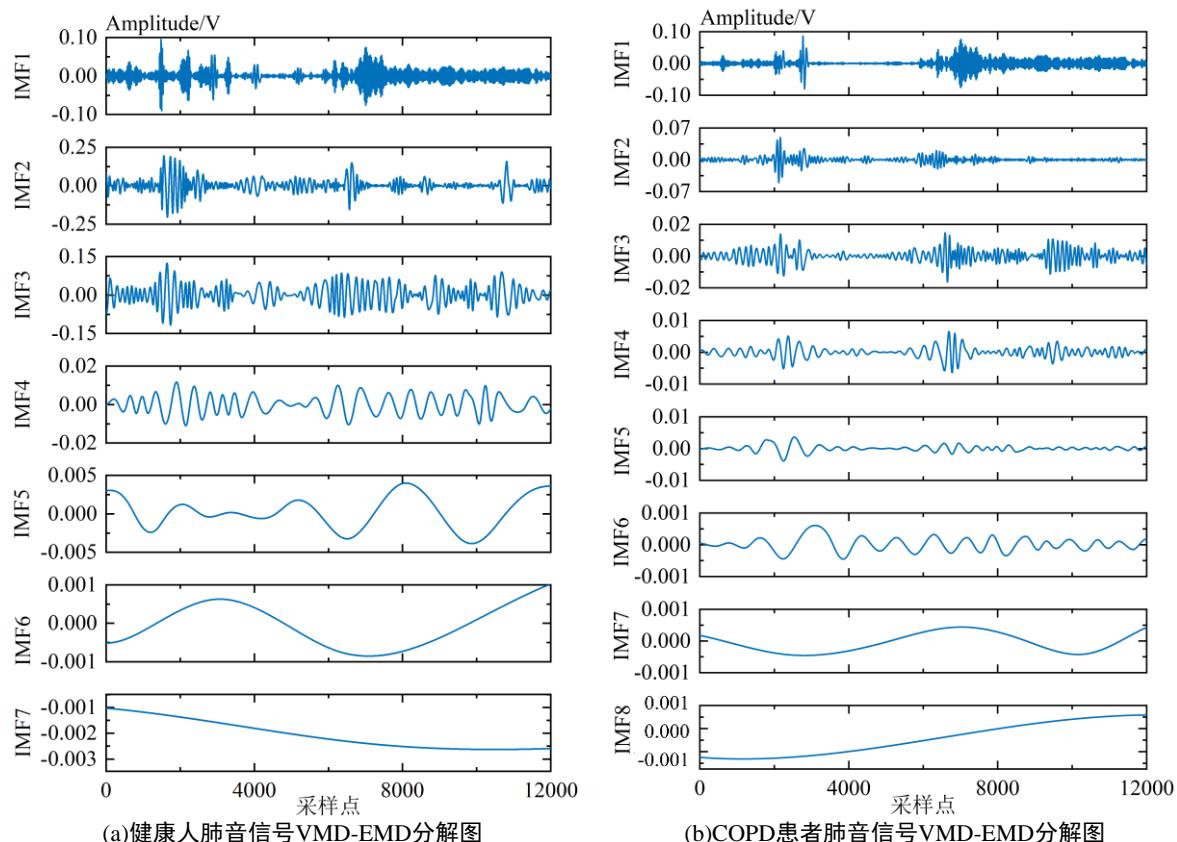


图 3-9 肺音信号 VMD-EMD 分解图

求取 EMD 分解的各个固有模态分量与原始信号的相关系数如表 3-4 所示。根据表 3-4 中的相关系数结果, 选取健康人的 IMF<sub>2</sub>、IMF<sub>3</sub>、IMF<sub>4</sub> 重构为 IMF<sub>c2</sub>, 最终将 IMF<sub>c2</sub> 与 IMF<sub>c1</sub> 重构为健康人肺音信号。同理, 选取 COPD 患者的 IMF<sub>3</sub>、IMF<sub>4</sub>、IMF<sub>5</sub> 重构为 IMF<sub>c2</sub>, 将 IMF<sub>c2</sub> 与 IMF<sub>c1</sub> 重构为 COPD 患者肺音信号, IVMD 分解后重构肺音信号波形图如图 3-10 所示。与图 3-6 相比, 信号波形更平滑, 毛刺更少, 信号强度更大, 说明 IVMD 算法去噪效果比前两种算法优异。

表 3-4 EMD 分解  $x_1(t)$  和  $x_1'(t)$  的各个固有模态分量相关系数

固有模态分量条数	健康人样本一	COPD 患者一
IMF <sub>1</sub>	0.0828	0.0792
IMF <sub>2</sub>	<b>0.3591</b>	0.1375
IMF <sub>3</sub>	<b>0.5319</b>	<b>0.4544</b>
IMF <sub>4</sub>	<b>0.6712</b>	<b>0.6135</b>
IMF <sub>5</sub>	0.1363	<b>0.3197</b>
IMF <sub>6</sub>	0.0632	0.1032
IMF <sub>7</sub>	0.1176	0.0390
IMF <sub>8</sub>	—	0.0014

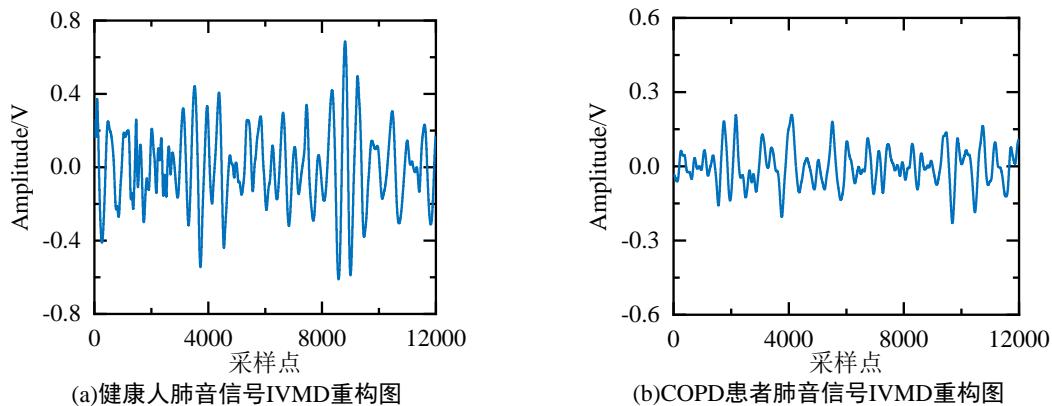


图 3-10 IVMD 分解后重构肺音信号波形图

### 3.5 实验结果与分析

#### 3.5.1 去噪评价指标

对去噪重构后的肺音信号进行量化评价，通常量化指标选用信噪比( $SNR$ )、均方误差和相关性。在信号处理方向中， $SNR$  是信号中的有效信息的功率与噪声的功率之比， $SNR$  的定义如式(3-1)所示。

$$SNR = 10\lg(P_y / P_c) \quad (3-1)$$

式中， $P_y$  表示信号中的有效信息的功率， $P_c$  表示原始信号中的噪声功率。

根据信噪比的含义可知，需要预先知道原始信号中的噪声功率和信号有效信息的功率。但是在实际工程中往往无法预先得知噪声功率和有效信息功率，因此本文对信噪比进行延伸为  $SNR'$ ，表示含噪声信号的功率与重构信号的功率之比，表达式如式(3-2)所示。

$$SNR' = 10 \lg(P_y' / P_c') \quad (3-2)$$

式中,  $P_y'$  表示原始信号的功率,  $P_c'$  表示重构信号的功率。  $SNR'$  的值越大表示去除的噪声功率越大, 去噪效果越好; 反之,  $SNR'$  的值越小表示去除的噪声功率越小, 去噪效果较差。均方误差表示信号处理前后的失真情况, 表达式如式(3-3)所示。

$$MSE = \frac{1}{n} \sum_{t=1}^n [c(t) - x(t)]^2 \quad (3-3)$$

式中,  $x(t)$  表示原始肺音信号,  $c(t)$  表示重构的肺音信号,  $n$  表示采样点,  $MSE$  的值越小表明去噪处理前后误差越小, 信号的失真情况较弱; 反之,  $MSE$  的值越大表明去噪处理前后误差越大, 信号的失真情况严重。

相关系数  $R$  表示重构信号与原始信号的相似程度, 其表达式如式(3-4)所示。

$$R = \frac{\sum_{t=0}^N x(t)c(t)}{\sqrt{\sum_{t=0}^N x^2(t) \sum_{t=0}^N c^2(t)}} \quad (3-4)$$

式中,  $x(t)$  表示含噪声降采样后的原始信号,  $c(t)$  表示重构后的去噪信号,  $N$  表示采样点。 $R$  的值越大, 表示重构信号与原始信号相关性越强, 重构信号可以更好地代表原始信号; 反之,  $R$  的值越小, 表示重构信号与原始信号相关性越弱, 重构信号不能充分地表征原始信号。

### 3.5.2 实验结果对比

将肺音信号经过两种不同的算法去噪后并进行重构,  $SNR'$ 、 $MSE$  和相关性结果对比值如表 3-5 所示, 去噪结果对比图如图 3-11 所示, 去噪波形图对比图如图 3-12 所示。

表 3-5 肺音信号去噪效果对比

评价指标	去噪方法	
	EMD	IVMD
健康人	$SNR'$	8.6554
	$MSE$	0.0064
	相关性	0.9308
COPD 患者	$SNR'$	7.3740
	$MSE$	0.0076
	相关性	0.9044

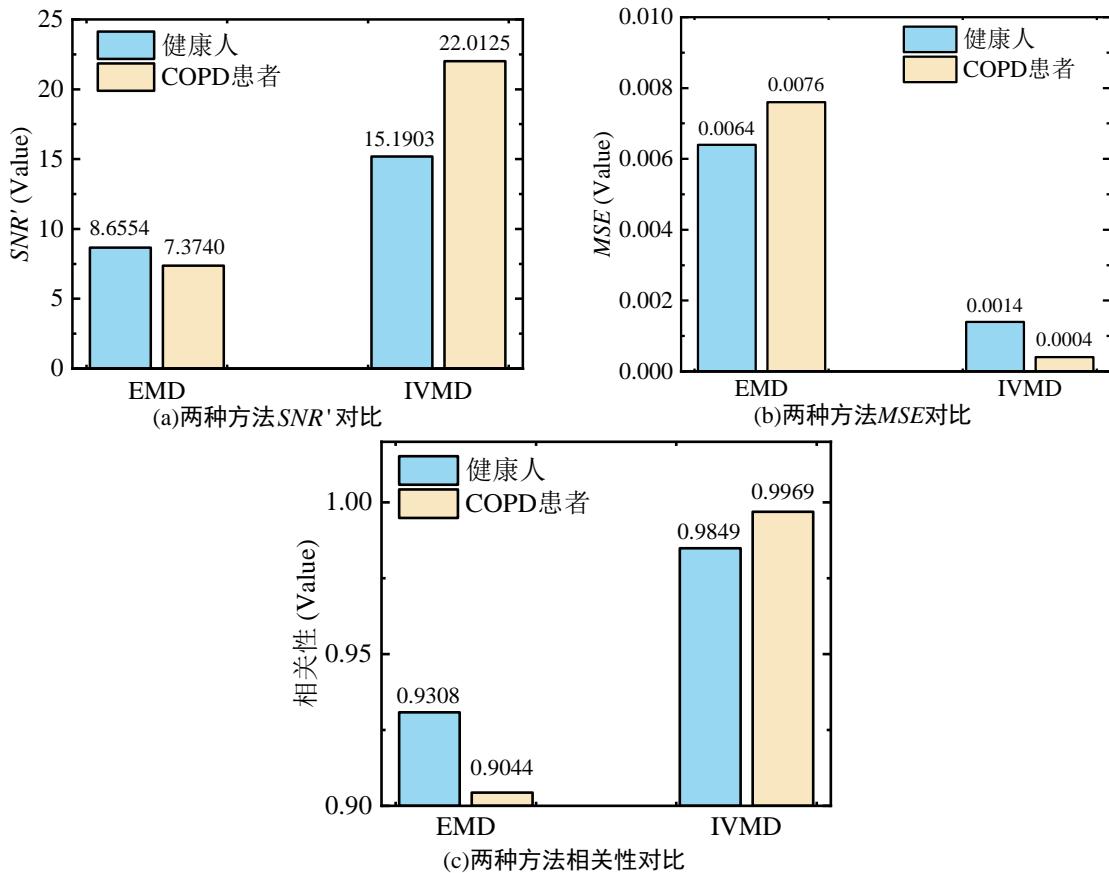


图 3-11 两种方法去噪结果对比图

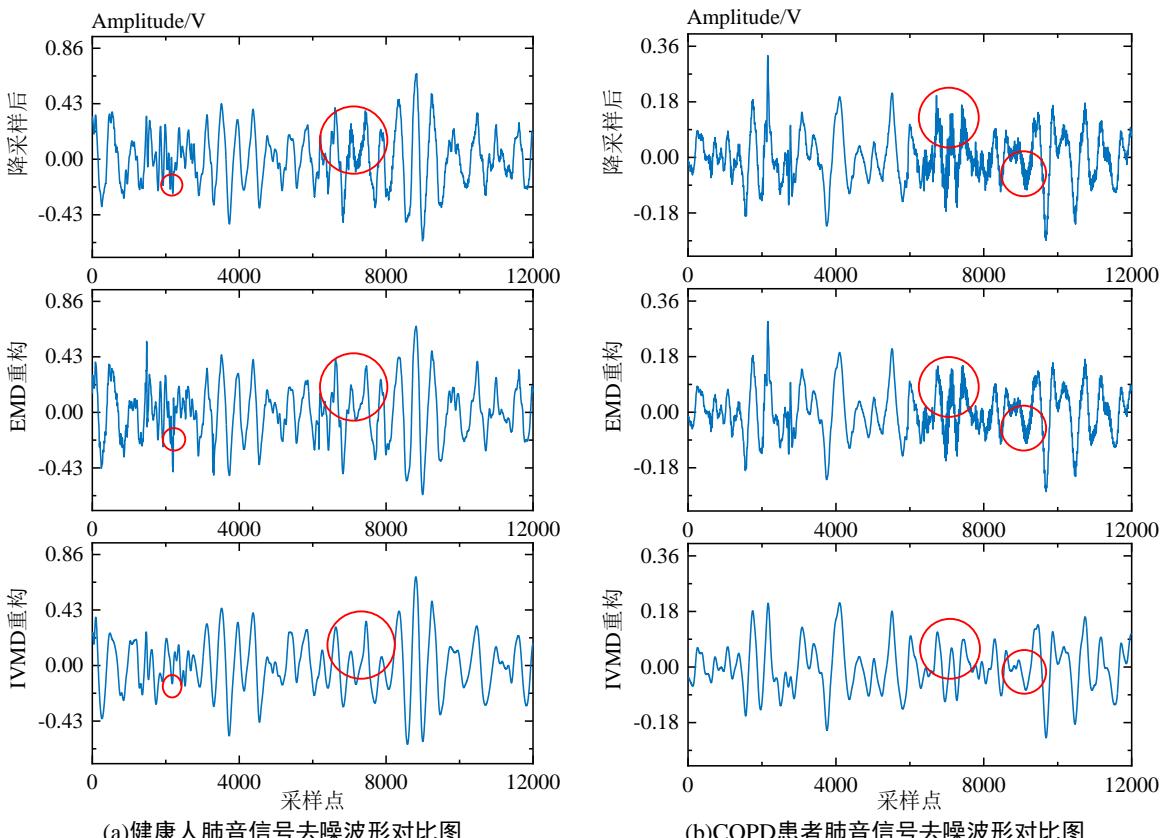


图 3-12 两种方法去噪波形对比图

由表 3-5 和图 3-11 可以得知, 对于健康人和 COPD 患者, IVMD 在两种算法中的去噪效果是最佳的。在  $SNR'$  方面, IVMD 比 EMD 提升效果显著。就  $MSE$  而言, 两种方法均取得优异效果, 其中 IVMD 的结果明显优于 EMD。在相关性方面, 结果表明基于相关系数的固有模态分量选取方法可行, 充分保留了原始肺音信号的特征, 且 IVMD 去噪效果优异并保留了信号的高相关性, 其中, 对于 COPD 患者的提升效果优于健康人的效果, 原因可能是 COPD 患者呼吸过程中伴随咳嗽或者气道受阻, 导致采集到的信号包含太多噪声, 从而使得去噪效果更加显著。由图 3-12 标记位置可得, IVMD 重构后的肺音波形毛刺明显减少, 波形更光滑, 更能反映肺音信号的特点, 说明 IVMD 去噪效果最优, 同时可以得出对 COPD 患者的去噪效果优于健康人的去噪效果。

### 3.6 本章小结

本章首先介绍了肺音信号采集系统的搭建。为了方便处理数据, 进行降采样和归一化采样点处理, 采用基于相关系数的经验模态分解对肺音信号进行自适应去噪, 为了抑制 EMD 的模态混叠, 采用具有充实的数学理论基础的 VMD 和 EMD 算法相结合, 解决信号的模态混叠现象, 实现对肺音信号的去噪处理。实验结果表明, 基于 IVMD 算法的肺音信号去噪算法效果更优, 有利于后续对肺音信号进行特征提取。

## 第四章 肺音信号的特征提取

上一章介绍了肺音信号的采集和预处理，去除了环境噪声和心音干扰，得到了纯净的肺音信号。本章介绍肺音信号的特征提取，根据肺音信号的特性，提取其时间序列倒谱系数以及梅尔频率倒谱系数，再将两种特征参数融合，方便后续的肺音信号分类。

### 4.1 肺音信号的线性预测倒谱系数

#### 4.1.1 线性预测系数

线性预测系数(Linear Prediction Coefficient, LPC)是求线性预测倒谱系数的理论基础。求 LPCC 的过程可概括为：首先对 LPC 进行快速傅里叶变换，然后对变换后的结果取对数，最后通过快速傅里叶逆变换求得 LPCC。

线性预测系数是建立一个数学模型模拟待分析的信号，即将待分析信号看做某一个数学模型的输出过程。该数学模型可以看作由激励模型、声道模型以及辐射模型构成，可以当作全极点参数模型进行参数估计，全极点参数模型的表达式如式(4-1)所示。

$$H(z) = \frac{X(z)}{U(z)} = \frac{G}{1 - \sum_{i=1}^N a_i z^{-i}} = \frac{G}{A(z)} \quad (4-1)$$

式中， $X(z)$  表示实际输入时间序列  $x(n)$  的 Z 变换， $U(z)$  表示输出信号  $z(n)$  的 Z 变换， $G$  是增益因子，作用是调节输出信号的幅值， $N$  是模型包含的极点阶数， $A(z)$  表示预测误差滤波器，也称为逆滤波器，其作用是对参数进行反向线性变换。

线性预测系数已经广泛应用于肺音信号，其中运用到的线性预测分析基本思路是：一个时间序列的某一时刻采样值可以利用该时间序列这一时刻之前的若干时刻采样值线性组合逼近。由于预测中存在误差，所以需要通过一组预测系数使得实际的信号采样值和线性预测的采样值误差最小，即令均方误差取得最小值。然后根据这一组线性预测系数有效精确地反映出信号的特征。

根据上述基本思路，对  $N$  阶肺音信号的线性预测分析可以表示为在某一时刻的采样值可以由该时刻之前的  $N$  个采样值线性组合进行预测，表达式如式(4-2)所示。

$$x(n) = \sum_{i=1}^N a_i x(n-i) \quad (4-2)$$

式中， $x(n)$  表示线性预测采样值。

实际采样值与线性预测采样值之间的误差计算表达式如式(4-3)所示。

$$\varepsilon(n) = x(n) - \hat{x}(n) = x(n) - \sum_{i=1}^N a_i x(n-i) \quad (4-3)$$

当上式误差取得最小值，预测采样值即为  $x(n)$  的最优估计，此时预测的均方误差值取得最小值。预测的均方误差值求解表达式如式(4-4)所示。

$$E_n = E[\varepsilon^2(n)] = E\{\{x(n) - \sum_{i=1}^N a_i x(n-i)\}^2\} \quad (4-4)$$

对上式中的线性预测系数  $a_i (i=1, 2, \dots, N)$  求偏导，将其结果设为 0，表达式如式(4-5)所示。

$$\frac{\partial E_n}{\partial a_i} = 0, \quad i = 1, 2, \dots, N \quad (4-5)$$

根据式(4-5)，可将式(4-4)化简为表达式如式(4-6)所示。

$$x(n-i)x(n) = \sum_{i=1}^N a_i x(n-i)x(n-j), \quad 1 \leq j \leq N \quad (4-6)$$

若定义表达式如式(4-7)所示。

$$\phi(i, j) = x(n-i)x(n-j) \quad (4-7)$$

则式(4-6)化简为表达式如式(4-8)所示。

$$\phi(i, 0) = \sum_{i=1}^N a_i \phi(i, j) \quad (4-8)$$

式(4-8)即为一个含有  $N$  个未知数的方程组，对其进行求解，可以得到预测系数  $a_i (i=1, 2, \dots, N)$ 。

采用莱文逊-杜宾算法计算线性预测系数，首先计算自相关系数  $r(j), j=1, \dots, N$ 。初始化  $i=0$ 。令  $E_0 = r(0), a_0 = 1$ 。根据以下公式进行递推计算：

$$k_i = \frac{r(i) - \sum_{j=1}^{i-1} a_j^{(i-1)} r(j-i)}{E_{i-1}} \quad (4-9)$$

$$a_i^{(i)} = k_i \quad (4-10)$$

$$a_j^{(i)} = a_j^{(i-1)} - k_i a_{i-j}^{(i-1)} \quad (4-11)$$

$$E_i = (1 - k_i^2) E_{i-1} \quad (4-12)$$

式中， $a_j^{(i)}$  表示系统中第  $i$  阶预测器的第  $j$  个预测系数， $E_i$  表示经过第  $i$  阶的预测器之后得到的预测误差。

当  $i > N$  时，递推结束，否则返回上述的递推公式继续递推。

### 4.1.2 线性预测倒谱系数

线性预测系数(LPC)已经广泛应用于语音信号中，但是其对误差非常敏感，小的误差也会严重影响频谱质量，而线性预测倒谱系数参数量少，计算速度更快，所以选择线性预测倒谱系数对肺音信号进行特征提取。

根据信号前  $p$  个时刻的采样值进行线性组合逼近某一时刻的采样值，具体过程如式(4-13)所示。

$$l_1(n) = \sum_{j=1}^p a_j s_H(n-j), (j=1, 2, 3, \dots, p) \quad (4-13)$$

式中， $l_1(n)$  表示  $n$  处的预测信号值， $a_j$  表示线性预测系数， $s_H(n)$  表示预加重后的信号。

通常取均方差来确定预测误差  $\varepsilon$ ，如式(4-14)所示。

$$\varepsilon^2 = \frac{1}{N} \sum_{n=0}^{N-1} [s_H(n) - l_1(n)]^2 \quad (4-14)$$

此时自相关系数如式(4-15)所示。

$$R(m) = \sum_{n=0}^{N-1} s_H(n) s_H(n+m) \quad (4-15)$$

联立上述三式对  $a_j$  求导，如式(4-16)所示。

$$R(j) = \sum_{i=0}^p a_i(n) R(j-i), (j=1, 2, 3, \dots, p) \quad (4-16)$$

根据 LPC 的参数可以计算得到同阶的 LPCC 参数，如式(4-17)所示。

$$C_{LPCC}(q) = C_{LPC}(q) + \sum_{j=1}^{q-1} \frac{q-j}{q} C_{LPCC}(q-j) C_{LPC}(q), (q=1, 2, 3, \dots, p) \quad (4-17)$$

式中， $C_{LPCC}(q)$  表示  $q$  阶的 LPCC。

根据线性预测倒谱系数的基本原理，可以得出线性预测倒谱系数可以很好地建立符合肺音信号的数学模型，更逼近肺音信号的时间特性，模型的参数相对简单，计算方便易于实现。

## 4.2 肺音信号的梅尔频率倒谱系数组提取

### 4.2.1 梅尔频率倒谱系数

在语音信号研究领域，梅尔频率倒谱系数已经得到充分的应用。根据生理学研究，人耳对不同频率的声音具有不同的听觉感知。当两个响度不等的声音同时传入人耳时，信号中的较高频率成分会影响到较低频率成分的听觉感知，使低频成分不易察觉，这种现象被称作掩蔽效应<sup>[64]</sup>。由于频率较低的声音在人耳中的传播距离大于频率较高的声音，所以一般来说，低音掩蔽高音。低

频声音掩蔽的临界带宽较高频要小。梅尔频率倒谱系数模拟人耳的非线性特性，利用一组滤波器模拟耳蜗的功能。梅尔频率与实际频率的关系如式(4-18)所示。

$$\text{Mel}(f) = 2595 \cdot \log\left(1 + \frac{f}{700}\right) \quad (4-18)$$

梅尔频率倒谱系数的算法流程如下：

步骤一：对原始信号进行预加重。预加重是为了提升信号的高频部分，弥补采集过程中高频成分的损失，通常采用高通滤波器。如式(4-19)所示。

$$H(z) = 1 - az^{-1} \quad (4-19)$$

式中， $a$  表示预加重系数，一般取值为 0.97。

步骤二：对预加重后的信号进行分帧。每  $N$  个采样点为一个单位称为一帧。为了避免相邻两帧差异较大，一般设置重叠区域  $M$  为帧长的  $1/2$ 。

步骤三：对每一帧信号进行加窗处理。汉明窗幅频特性是旁瓣衰减较大，可以增加相邻帧之间的连续性，泄露误差较小，汉明窗的表达式如式(4-20)所示。

$$W(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), 0 \leq n \leq N-1 \quad (4-20)$$

步骤四：对加窗后的每一帧信号进行快速傅里叶变换，将信号转换到频域上，得到每一帧的频谱，并对频谱取模平方得到信号的功率谱，快速傅里叶变换如式(4-21)所示。

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi nk/N}, 0 \leq n, k \leq N-1 \quad (4-21)$$

式中， $x(n)$  表示输入信号， $N$  表示采样点数。

步骤五：将快速傅里叶变换后的功率谱通过 Mel 滤波器组进行滤波和刻度变换。Mel 滤波器组是一组重叠的三角带通滤波器组，三角带通滤波器可以平滑频谱，消除谐波。每个滤波器的输出响应如式(4-22)所示。

$$H_m(k) = \begin{cases} 0 & , k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)} & , f(m-1) \leq k < f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)} & , f(m) \leq k < f(m+1) \\ 0 & , f(m+1) \leq k \end{cases} \quad (4-22)$$

式中， $k$  表示信号频率， $f(m-1)$ 、 $f(m)$ 、 $f(m+1)$  依次表示第  $m-1$ 、 $m$ 、 $m+1$  个滤波器的中心频率。

步骤六：对每个滤波器滤波后的信号取对数能量。取对数能量如式(4-23)所示。

$$E_m = \ln\left(\sum_{k=0}^{N-1} |X(k)|^2 H_m(k)\right), 0 \leq m \leq M \quad (4-23)$$

式中,  $|X(k)|^2$  表示每一帧信号的能量频谱,  $H_m(k)$  表示频率响应,  $M$  表示滤波器的个数。

步骤七: 将对数能量进行离散余弦变换。可将滤波器的能量矢量正交化, 去除冗余信息。MFCC 特征参数表达式如式(4-24)所示。

$$C_n = \sum_{k=0}^{N-1} E_m \cos\left(\frac{\pi n(m-0.5)}{M}\right), n=1, 2, 3, \dots, L \quad (4-24)$$

式中,  $L$  表示 MFCC 系数的阶数,  $C_n$  表示 MFCC 的特征参数。

#### 4.2.2 梅尔频率倒谱系数提取

根据对 MFCC 参数提取的基本原理可知, MFCC 特征参数提取的流程图如图 4-1 所示。

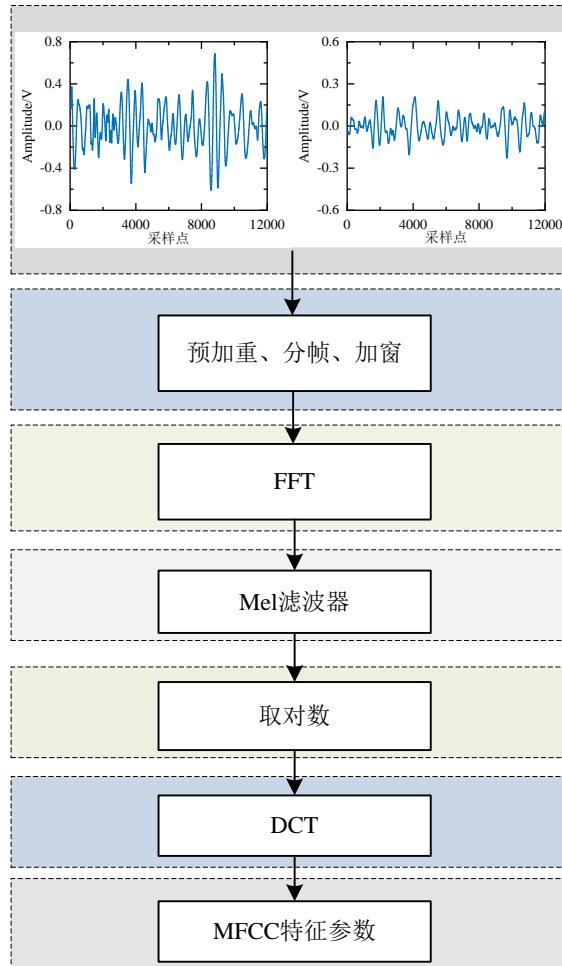


图 4-1 MFCC 特征参数提取流程图

在实际求取 MFCC 特征参数过程中, 具体求解过程如下所示:

步骤一: 首先对第三章去噪重构后的肺音信号进行采样和量化, 并对其进行预加重、分帧和加窗等预处理, 得到每一帧的肺音信号序列  $x(n)$ 。

步骤二：对步骤一预处理得到的肺音信号序列  $x(n)$  进行 FFT，将肺音信号变换到频域上，得到频谱  $X(n)$ ，再对  $X(n)$  求取幅度谱  $|X(n)|^2$ 。

步骤三：将步骤二得到的幅度谱  $|X(n)|^2$  经过 Mel 滤波器组幅度谱  $M_m(k)$  得到参数  $P_m$ ，其表达式如式(4-25)所示。

$$P_m = \sum_{k=f_{m-1}}^{f_m} H_m(k)S_n(k), m=1,2,\dots,M \quad (4-25)$$

步骤四：对参数  $P_m$  进行取对数能量操作，得到参数  $L_m$ ，再对其进行 DCT 变换，将参数  $L_m$  变换到倒频域中，从而得到了梅尔频率倒谱系数  $C_{MFCC}$ 。

通过对梅尔频率倒谱系数的详细阐述，可以得知：梅尔频率倒谱系数是非线性的，符合人耳的听觉感知特性，可以加强肺音信号的低频细节，同时不需要建立数学模型，适用范围广泛，而且在求解过程中对肺音信号进行频率转换，得到了更丰富的肺音信号特征。

### 4.3 肺音信号的 LPCC-MFCC 特征参数融合

由上述对 LPCC 的阐述可知，LPCC 可以表征肺音信号的频谱幅度特征，且计算量小、易于实现。但是 LPCC 是基于全极点数字参数模型提取的特征，在频率上都是线性的，并没有考虑实际人耳的听觉感知特性，不能够完全反映肺音信号的非线性特性，因此会导致后续分类模型的性能下降。与 LPCC 相比，MFCC 是非线性的，可以更好地反映实际人耳的听觉感知特性，具有更良好的特征属性。但是根据上述 MFCC 的原理可知，MFCC 中的 Mel 滤波器在高频区域内，滤波器分布稀疏且数量较少，这就会导致 MFCC 的计算精度在高频率处会下降，无法更精细地反映高频肺音信号的特征。此外，在实际的应用环境中，肺音信号包含大量环境噪声，会对 MFCC 造成干扰。因此单一的特征参数并不能充分提取肺音信号的特征，会影响分类的准确率。针对此问题，本文将 LPCC 和 MFCC 两种特征参数融合形成新的特征参数，融合特征参数可以具有更多维度肺音信号的特征，从而表现出更好的性能。融合的步骤如下所示：

步骤一：首先计算 LPCC 特征参数  $C_{LPCC}$ ，其表达式如式(4-26)所示。

$$C_{LPCC} = \begin{bmatrix} L_{11} & L_{12} & \cdots & L_{1L} \\ L_{21} & L_{22} & \cdots & L_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ L_{N1} & L_{N2} & \cdots & L_{NL} \end{bmatrix} \quad (4-26)$$

式中， $L_{NL}$  表示第  $N$  帧，第  $L$  维 LPCC 特征参数， $C_{LPCC}$  为  $N$  行  $L$  列的矩阵。

步骤二：计算 MFCC 特征参数  $C_{MFCC}$ ，其表达式如式(4-27)所示。

$$C_{\text{MFCC}} = \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1M} \\ M_{21} & M_{22} & \cdots & M_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ M_{N1} & M_{N2} & \cdots & M_{NM} \end{bmatrix} \quad (4-27)$$

式中,  $M_{NM}$  表示第  $N$  帧, 第  $M$  维 MFCC 特征参数,  $C_{\text{MFCC}}$  为  $N$  行  $M$  列的矩阵。

步骤四: 将 MFCC 特征参数  $C_{\text{MFCC}}$  连接在 LPCC 特征参数  $C_{\text{LPCC}}$  后面进行特征融合, 构成融合特征向量  $C$ , 其表达式如式(4-28)所示。

$$C = \begin{bmatrix} L_{11} & L_{12} & \cdots & L_{1L} & M_{11} & M_{12} & \cdots & M_{1M} \\ L_{21} & L_{22} & \cdots & L_{2L} & M_{21} & M_{22} & \cdots & M_{2M} \\ \vdots & \vdots \\ L_{N1} & L_{N2} & \cdots & L_{NL} & M_{N1} & M_{N2} & \cdots & M_{NM} \end{bmatrix} \quad (4-28)$$

本文设置帧长为 25ms, 帧移为 15ms, 加汉明窗,  $L=12$ ,  $M=13$ , 对每一帧信号求 LPCC 特征参数和 MFCC 特征参数, 即融合特征为 25 维的高维度特征向量。融合特征 LPCC-MFCC 包含有肺音信号更多的特征信息, 既能够充分提取肺音信号的特征, 又适用于后续深度学习模型训练的大样本量的要求。融合特征 LPCC-MFCC 三维图如图 4-2 所示。

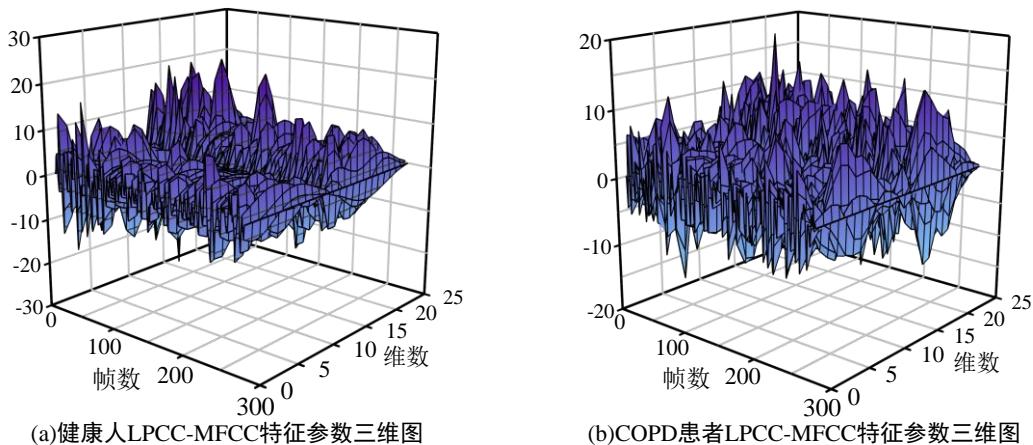


图 4-2 LPCC-MFCC 特征参数三维图

图 4-2(a)表示健康人 LPCC-MFCC 特征三维图, 图 4-2(b)表示 COPD 患者 LPCC-MFCC 特征三维图。其中前 12 维来自 LPCC 特征参数, 后 13 维来自 MFCC 特征参数, 由图 4-2 可得, 两者存在差异性, 可以作为后续的分类依据。

#### 4.4 本章小结

本章将肺音信号的 LPCC 特征参数和 MFCC 特征参数进行融合。首先对线性预测系数和梅尔频率倒谱系数的基本原理与计算过程进行了介绍, 然后根据肺音信号的特性提取两种特征参数, 最后将两种特征参数进行融合, 得到新的特征参数, 为后续分类提供基础。

## 第五章 基于 CNN-Transformer 的肺音信号分类模型研究

经过第三章降采样、去噪等预处理操作和第四章的肺音信号特征提取与融合特征之后，原始的肺音信号已经被转换为可以表征肺音信号本质的特征参数。卷积神经网络模型具有对输入的信号自动地提取特征并且具有权值共享的特点，本章根据上述特点选择 CNN 为主体框架对肺音信号进行客观分析，并针对卷积神经网络中 Softmax 函数泛化能力差的问题，采用泛化能力强的支持向量机进行改进；为了提取肺音信号的全局特征，采用 CNN-Transformer 网络模型，提升分类网络模型的性能。

### 5.1 基于 CNN 的肺音信号分类模型研究

卷积神经网络具有深层的前馈神经网络，方便进行卷积计算。CNN 作为深度学习模型的经典模型之一，基本思路是先通过卷积层提取信号的特征，然后通过池化层选取更有利于分类的特征。

本文采用常用于时间序列信号分类处理的卷积神经网络。本小节利用卷积神经网络为框架搭建肺音信号分类模型，其模型结构如图 5-1 所示。

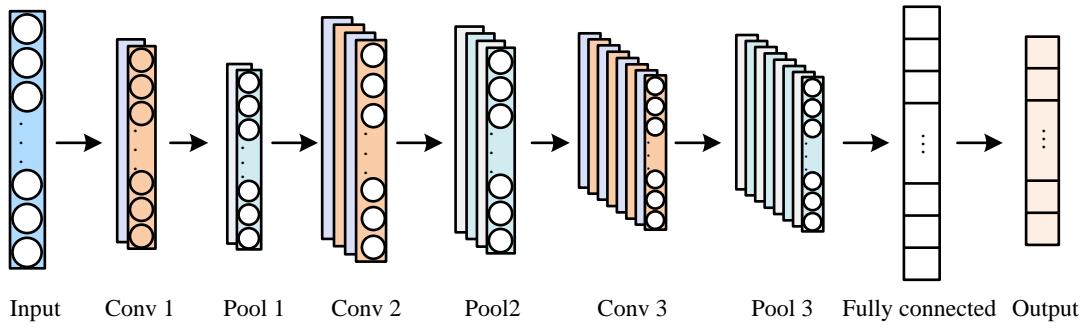


图 5-1 卷积神经网络模型示意图

根据上述构建的分类模型，对肺音信号进行分类的具体训练流程如下所示：

- (1) 模型数据准备阶段，首先将经过去噪和特征提取后的健康人和 COPD 患者两种人群的融合特征参数按照 8:2 的比例划分为训练集和测试集。再对训练集随机打乱顺序，训练分类模型。
- (2) 将训练集中的融合特征参数作为卷积神经网络的输入，设定网络为二分类，并对网络模型参数进行初始化，设置权重和偏置量，开始训练网络模型。
- (3) 根据肺音信号融合特征本身的特点，设置卷积层、池化层以及步长，使用 Relu 激活函数对每一层卷积层进行激活，池化方式选择平均池化，对输入信号进行降低参数量的同时保留肺音信号的更多特征。对卷积神经网络的输入进行遍历，充分提取肺音信号的深层特征。
- (4) 为了防止网络模型过拟合，通过加入 Dropout 层可以丢弃部分内部神经元的连接，达到减少模型计算量的目的。卷积层提取深层肺音信号特征后，通过全连接层进行融合，最后经过

Softmax 层输出肺音信号的分类结果。

(5)根据训练的结果，保存训练结果最优的网络模型，再应用到测试集进行验证，得到最终的分类结果。利用交叉熵损失函数以及神经网络的反向传播函数对网络模型的参数进行更新，最终实现对肺音信号深层次的特征提取与分类。卷积神经网络在第四章提取融合特征的基础上，对肺音信号进行更深层次的特征提取，利用权值共享特性降低模型的复杂度。

## 5.2 基于 CNN-SVM 肺音信号分类模型研究

### 5.2.1 支持向量机

支持向量机是一种按照监督学习方式运作的二分类线性分类器，在特征空间上间隔最大。支持向量机将原问题转化为凸优化问题，再将凸优化问题转为二次规划问题并求解。由于支持向量机泛化能力强，因此本文利用支持向量机进行分类，不仅可以提高训练速度，也可以提升分类精度以及泛化能力。

支持向量机的求解分为两个部分：第一部分是将原问题转化为凸优化问题，第二部分是对第一部分转化后的凸优化问题进行求解。

(1) 将原问题转化为凸优化问题。

步骤一：若存在一个超平面将样本分开，使得正负样本之间的间隔最大，则该超平面表达式如式(5-1)所示。

$$f(x) = w^T x + b \quad (5-1)$$

式中， $b$  表示位移项，决定超平面与原点的距离， $w$  表示平面的法向量，决定超平面的方向。

当上式等于 0 时， $x$  即在超平面上，超平面示意图如图 5-2 所示。

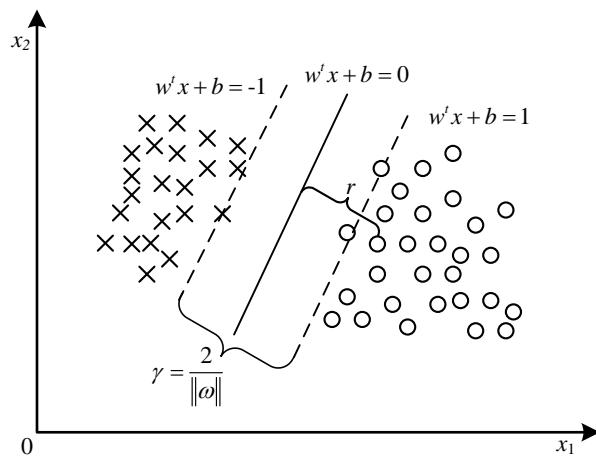


图 5-2 超平面示意图

步骤二：设置决策边界。若存在一个样本可令  $w^t x + b = 1$  和  $w^t x + b = -1$  成立且距  $w^t x + b = 0$  最近的超平面即为所求得支持向量。假设两个不同类别的支持向量到  $w^t x + b = 0$  的间隔之和用  $\gamma$  表示，则  $\gamma$  最大时即为最优解，因此原问题转化为凸优化的二次规划问题，构建原始模型如式(5-2)所示。

$$\min_{w,b} \frac{1}{2} \|w\|^2, s.t. y_t (w^t x + b) \geq 1, (t=1,2,\dots,m) \quad (5-2)$$

(2)对凸优化问题进行求解。

步骤一：通过计算，当约束条件为  $w^t x + b \geq 1$  时，两个平面的距离是  $\gamma = \frac{2}{\|w\|}$ 。此时问题转

化为如何使得这两个平面之间的距离最小。利用拉格朗日乘子法，可将两个参数的优化问题转化为一个参数的优化问题，进而求解，得到拉格朗日函数表达式如式(5-3)所示。

$$L(w,b,\alpha) = \frac{1}{2} \|w\|^2 + \sum_{t=1}^l \alpha_t [1 - y_t (w^t x_t + b)] \quad (5-3)$$

式中， $\alpha_t = (\alpha_1, \alpha_2, \dots, \alpha_m)$  为拉格朗日算子。令  $L(w,b,\alpha)$  对  $w$  和  $b$  偏导为零，可得式(5-4)。

$$\begin{cases} w = \sum_{t=1}^m \alpha_t y_t x_t \\ 0 = \sum_{t=1}^m \alpha_t y_t \end{cases} \quad (5-4)$$

将式(5-4)代入式(5-3)可得表达式如式(5-5)所示。

$$\max_{\alpha} -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) + \sum_{j=1}^l \alpha_j \quad (5-5)$$

若存在约束条件如式(5-6)所示。

$$\begin{cases} \alpha_t [y_t f(x_t) - 1] = 0 \\ y_t f(x_t) - 1 \geq 0 \\ \alpha_t \geq 0 \end{cases} \quad (5-6)$$

将式(5-6)代入(5-5)继续求解  $w, b, \alpha$ ，得到模型表达式如式(5-7)所示。

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^n y_i \alpha_i (x_i \cdot x) + b \right\} \quad (5-7)$$

式中， $x_i$  表示支持向量， $\text{sgn}$  为符号函数。

### 5.2.2 基于 CNN-SVM 的肺音信号分类模型

卷积神经网络的全连接层的作用是整合卷积层、池化层中用作区分类别的特征，从而尽可能

多地提取肺音信号的有效特征，最终实现肺音信号的分类。CNN-SVM 网络模型首先舍弃原始网络的全连接层和 Softmax 分类函数，将最后一层输出合并成特征向量，作为 SVM 层的输入。

基于 CNN-SVM 网络模型的肺音信号分类具体步骤如下所示：

- (1) 将两种人群的肺音信号按照 8:2 的比例划分为训练集和测试集。将重构肺音信号的融合特征作为网络模型的输入，训练分类模型。
- (2) 在该网络模型的卷积层后使用 Relu 激活函数，池化层选择平均池化方式。
- (3) 使用 SVM 层替换原始网络的 Softmax 分类函数。对于任意输入的值，输出为 0 到 1 的连续值，若输出值在 0.5 之上时分类为正样本，反之则分类为负样本。
- (4) 根据训练情况进行参数更新并保存最优模型，进行测试，输出最终分类结果。其网络示意图如图 5-3 所示。

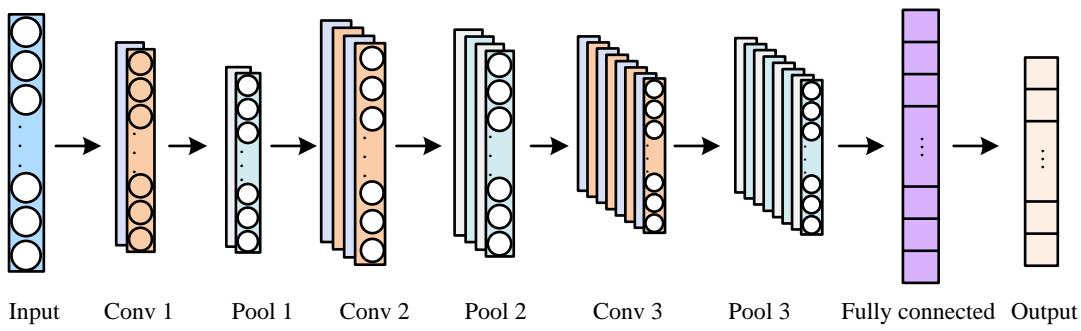


图 5-3 CNN-SVM 网络模型示意图

### 5.3 基于 CNN-Transformer 的肺音信号分类模型研究

针对 CNN 对长时间序列无法有效提取特征的问题，本文采用 CNN-Transformer 的肺音信号分类模型，该模型主要由两层 CNN 网络和三层 Transformer 的编解码层组成，CNN-Transformer 网络模型示意图如图 5-4 所示。

该网络模型将第四章的融合特征作为输入，通过卷积层和池化层提取肺音信号的细节特征，组成时间序列向量矩阵，对时间序列向量矩阵进行位置编码并保存其绝对位置。然后经过编码层部分，首先利用多头注意力使得模型共同关注时间序列不同位置的信息，捕获肺音信号时间序列的内部结构特征，然后通过模型中的残差模块可以舍弃冗余层，防止网络退化，再对数据进行层归一化，加快网络的训练速度，提升网络模型的稳定性，最后利用前向反馈神经将输入信号映射到更高维度上。通过解码层的掩码多头注意力得到之前已经预测的输出信息，即记录当前时刻的输入信息。第二个多头注意力可以利用当前输入的信息得到下一时刻的信息，利用当前的输入与经过编码提取过的特征向量之间的关系来预测输出。最后经过线性变换和 SVM 得到分类结果。

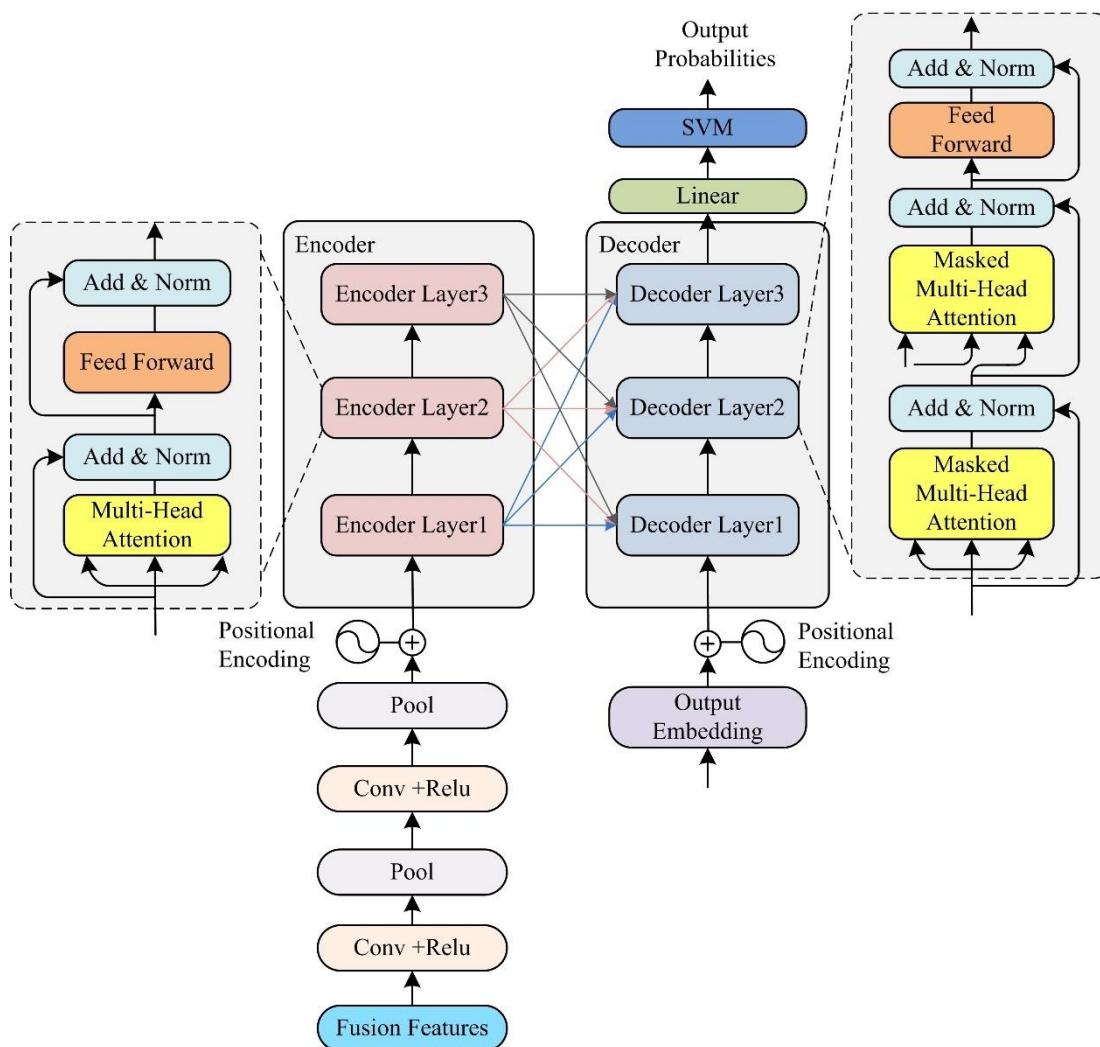


图 5-4 CNN-Transformer 网络模型示意图

## 5.4 实验结果与分析

### 5.4.1 实验数据及参数设置

实验数据来自 ICBHI 公有数据集的 500 组 COPD 患者数据<sup>[65]</sup>和实验室自采的 500 组健康人数据。ICBHI 公有数据集已被广泛引用，每条样本由专家注释。

本实验选择 Intel(R) Core(TM) i9-9900 K 的 CPU，参数为 8 核 16 线程，主频为 3.6GHz 显卡。本实验基于 Ubuntu 18.04.3 系统平台和 python3.8+Pytorch 框架，采用 Adam 优化器进行模型优化，设置初始学习率为 0.005，根据模型优化，最终学习率为 0.0001，迭代次数设置为 1500 次，batch size 设置为 64。

### 5.4.2 评价指标

混淆矩阵是深度学习中常用的一种态势分析表，主要用于总结模型的分类结果。在深度学习

中，通常使用通过真阳性(True Positive, TP)、真阴性(True Negative, TN)、假阳性(False Positive, FP)与假阴性(False Negative, FN)这 4 个通用的评价指标来更为清晰地对混淆矩阵进行表示，其中，TP 表示正样本被正确分类为正样本的个数；TN 表示负样本被正确分类为负样本的个数；FP 表示负样本被误分类为正样本的个数；FN 表示正样本被误分类为负样本的个数。

本研究通过以下四个评价指标来评定网络模型分类性能的优劣：准确率(Accuracy, Acc)、灵敏度(Sensitivity, Sen)、精确度(Precision, Pre)和 F1-score，其中精确度也被称为召回率。在肺音信号分类模型中，Acc 是指全部正确识别的肺音样本占全部肺音样本的比率，Acc 越高，表明该模型的检测性能越好；精确度是指某一类型肺音信号正确识别的数目占该类型所有数目的比率；Sen 是对目标做出正确的判断，Sen 的数值越高，则意味着目标鉴定越容易，即系统对于目标数据的筛查率越高；Pre 是指对假目标做出阳性反应的程度，Pre 越高，则误诊率越低，针对性更强。

(1)准确率表示用于计算正确样本量和样本总量的比值，如式(5-8)所示。

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (5-8)$$

(2)灵敏度也被称为召回率，表示的是所有正确样本中被分对的比例，衡量的是分类器的性能，如式(5-9)所示。

$$Sen = \frac{TP}{TP + FN} \quad (5-9)$$

(3)精确度表示正样本被分对的比例，衡量了分类器对正样本的分类性能，如式(5-10)所示。

$$Pre = \frac{TP}{TP + FP} \quad (5-10)$$

(4)综合评价指标 F1-score 是精确度和召回率的调和平均指标，尽量缩小两者差异，反映出模型对每种类别识别的综合性能，如式(5-11)所示。

$$F1 = 2 * \frac{\frac{TP}{TP + FP} * \frac{TP}{TP + FN}}{\frac{TP}{TP + FP} + \frac{TP}{TP + FN}} \quad (5-11)$$

#### 5.4.3 分类结果与分析

为了验证 CNN-Transformer 的有效性，将其与 CNN 和 CNN-SVM 进行对比，准确率和损失率迭代曲线对比图如图 5-5 所示。其中，图 5-5(a)表示准确率迭代曲线对比图，三网络模型的准确率均呈上升趋势，CNN 在迭代 540 次后逐渐收敛，准确率稳定在 0.83 左右，CNN-SVM 在迭代 320 次后逐渐收敛，准确率稳定在 0.88 左右，CNN-Transformer 在迭代 140 次后趋于稳定，准确率稳定在 0.95 左右。图 5-5(b)表示损失率迭代曲线对比图，CNN 在迭代 400 次后逐渐收敛，损

失率稳定在 0.32 左右, CNN-SVM 在迭代 240 次后逐渐收敛, 损失率稳定在 0.21 左右, CNN-Transformer 在迭代 120 次趋于稳定, 损失率稳定在 0.10 左右, 说明 CNN-Transformer 比另外两种网络收敛速度更快, 准确率更高。

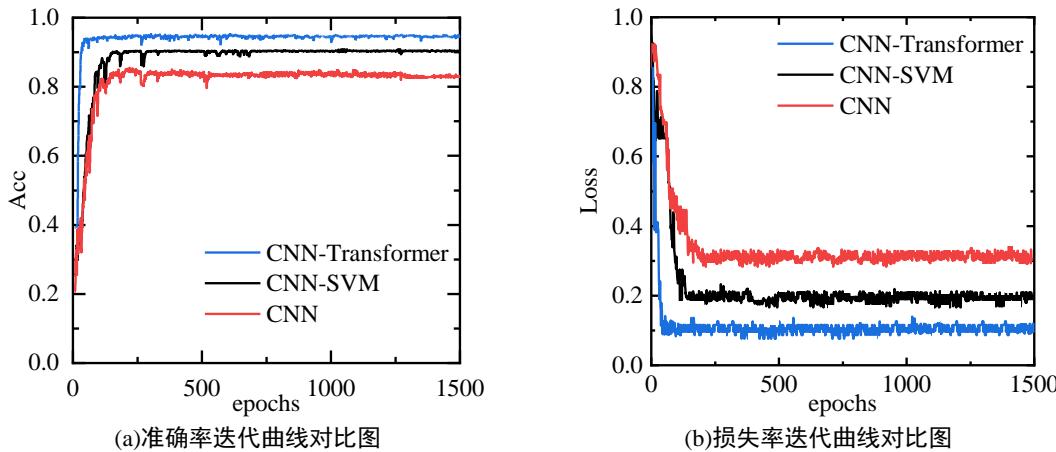


图 5-5 三种网络模型迭代曲线对比图

实验结果表明, 对肺音信号进行预处理并融合特征, 有效提升肺音信号分类的结果。本文采用的 CNN-SVM 和 CNN-Transformer 网络模型结果优于其他网络模型, 其中 CNN-Transformer 网络模型的准确率、灵敏度、精确度和 F1-score 分别为 95.70%、94.62%、93.87%、93.88%, 明显优于其他网络模型。原因是对肺音信号进行去噪后, 排除了噪声的干扰, 融合的特征更能表征肺音信号, 将卷积神经网络和 Transformer 的位置编码相结合, 有效地匹配肺音信号的时间序列特性, 充分提取肺音信号的时间序列深层特征, 使得网络模型分类更加准确, 网络性能更稳定。

为了进一步验证本文所提特征和所提模型在肺音信号分类方面的优越性, 将三种特征分别作为三种模型的输入, 通过模型的分类准确率、灵敏度、精确度、F1-score 进行评估, 三种模型的性能对比结果如表 5-1 所示, 三种模型中分类指标对比图如图 5-6 所示。

表 5-1 三种模型的性能对比结果

Models	特征	Acc	Sen	Pre	F1-score
模型一(CNN)	LPCC	76.90%	82.08%	81.39%	78.67%
	MFCC	79.35%	86.49%	83.64%	79.94%
	LPCC-MFCC	80.26%	89.75%	85.37%	82.46%
模型二(CNN-SVM)	LPCC	79.93%	85.37%	83.19%	79.16%
	MFCC	84.75%	87.54%	85.83%	80.28%
	LPCC-MFCC	88.38%	92.13%	88.65%	86.37%
模型三(CNN-Transformer)	LPCC	82.56%	89.75%	88.27%	87.94%
	MFCC	86.75%	90.34%	90.26%	89.53%
	LPCC-MFCC	<b>95.70%</b>	<b>94.62%</b>	<b>93.87%</b>	<b>93.88%</b>

由表 5-1 的结果可知, 融合特征 LPCC-MFCC 包含更多肺音信号的信息, 在 CNN-Transformer 网络模型中, 因加入了 SVM 和 Transformer, 提取了肺音信号的全局特征, 提升了网络模型的泛化性, 最终结果优于其他两个网络模型。

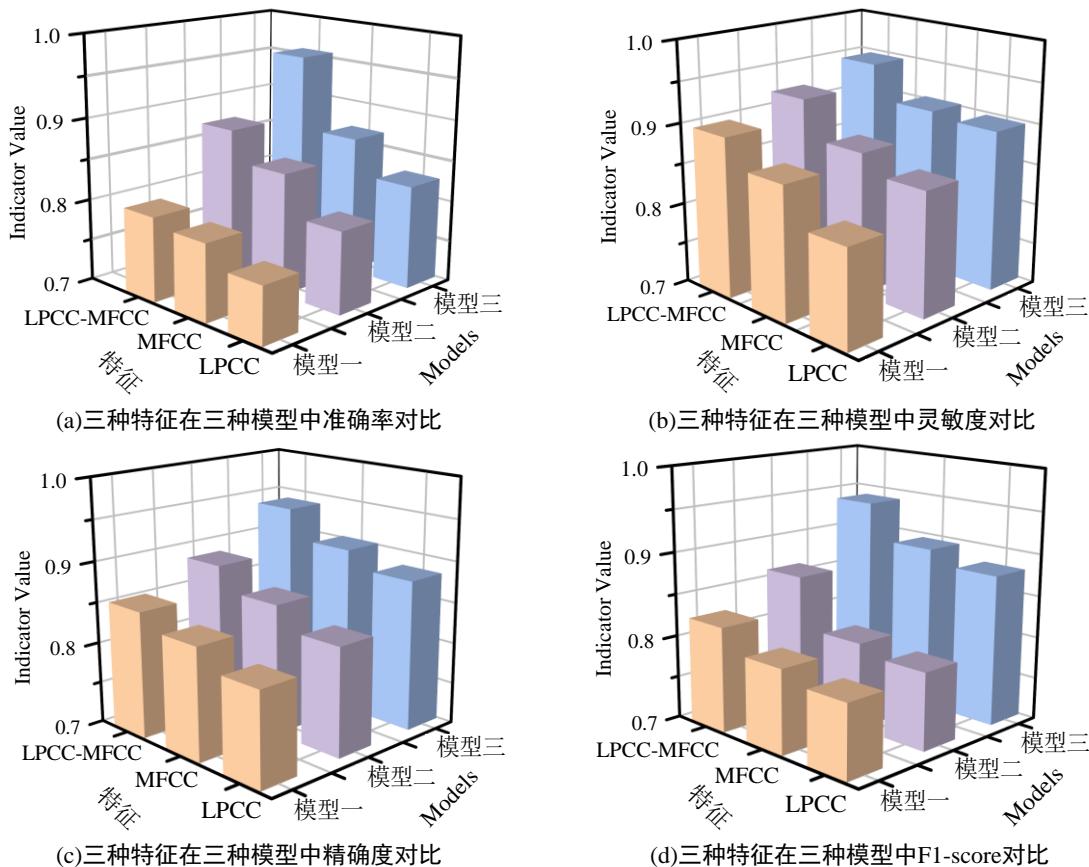


图 5-6 三种特征在三种模型中分类指标对比图

由图 5-6 可知, 在 CNN-Transformer 网络模型中, MFCC 与 LPCC 准确率相差 4%左右、灵敏度和精确度相差 2%左右、F1-score 相差 1%左右, 说明 MFCC 和 LPCC 均可以体现肺音信号的特性; 而融合特征 LPCC-MFCC 的准确率比单一特征提升 13%左右, 灵敏度和精确度提升 5%左右, F1-score 提升 6%左右, 说明融合特征包含了肺音信号的时间序列特征和非线性非平稳特征, 比单一特征更全面。特征为融合特征 LPCC-MFCC 时, CNN-Transformer 的准确率比 CNN 网络模型提升 15%左右, 比 CNN-SVM 网络模型提升 7%左右; 灵敏度比 CNN 网络模型提升 4%左右, 比 CNN-SVM 网络模型提升 2%左右; 精确度比 CNN 网络模型提升 8%左右, 比 CNN-SVM 网络模型提升 5%左右; F1-score 比 CNN 网络模型提升 11%左右, 比 CNN-SVM 网络模型提升 7%左右; 说明 CNN-Transformer 网络模型的各项性能优于其他两种模型。

为了证明本文所提 CNN-Transformer 网络模型在肺音信号分类方面的性能优异, 与在 ICBHI 公共数据集进行研究的其他文献进行对比, 对比结果如表 5-2 所示, 柱状图如图 5-7 所示。

表 5-2 与其他研究的分类结果对比

Models	Acc	Sen	Pre	F1-score
[66]	71.15%	83.00%	72.00%	77.00%
[67]	76.39%	52.78%	84.26%	68.52%
CNN-Transformer	<b>95.70%</b>	<b>94.62%</b>	<b>93.87%</b>	<b>93.88%</b>

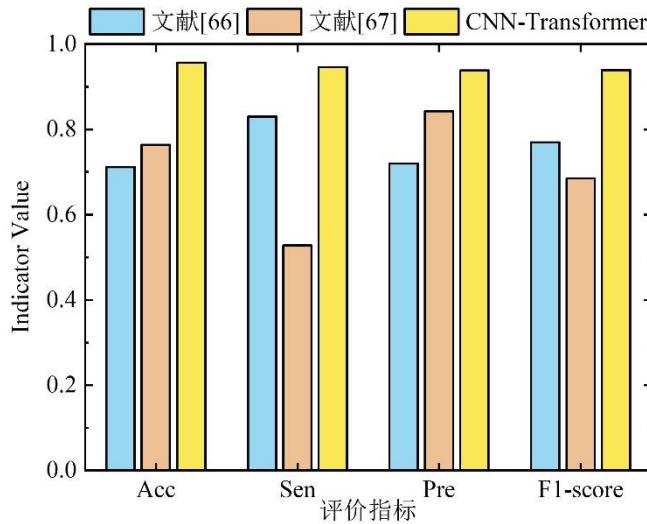


图 5-7 与其他研究的分类结果对比图

文献<sup>[66]</sup>通过卷积神经网络提取肺音信号特征，再利用线性判别和随机子空间对肺音信号进行分类。文献<sup>[67]</sup>提取肺音信号的频谱图，利用卷积神经网络结合长短时记忆神经网络对肺音信号进行了分类。由表 5-2 和图 5-7 可知，本文所提 CNN-Transformer 模型的分类准确率提升了 24% 左右，灵敏度提升了 21% 左右，精确度提升了 21% 左右，F1-score 提升了 16% 左右，充分表明本文所提融合模型和网络模型的优越性能，从而验证根据肺音信号自身的时间序列特性和非线性非平稳特性，进行特征提取和搭建模型可行性。

## 5.5 本章小结

本章采用了一种 CNN 与 Transformer 相结合的肺音信号分类模型。首先，选择轻量化的 CNN 作为网络模型的主体框架，利用泛化能力强的 SVM 代替全连接层，提升网络模型的泛化能力，提升网络模型的分类准确度。然后，为了进一步增强网络模型提取肺音信号的细节能力，引入 Transformer，利用其并行结构和特殊的位置编码，精确地提取肺音信号的时间序列特征。实验结果表明，相比于其他几种网络模型，本文采用的 CNN-Transformer 网络模型的准确率、召回率、精确度和 F1-score 均有所提升，可以准确地提取肺音信号的特征并进行分类，具有较好的泛化性和稳定性，融合特征比单一特征更能反映肺音信号的特征，有助于提升网络模型的准确率，最终融合特征在 CNN-Transformer 网络模型的准确率比 CNN-SVM 提升了 7.32%，召回率提升了 2.49%，精确度提升了 5.22%，F1-score 提升了 1.51%。

## 第六章 总结与展望

### 6.1 总结

随着复杂环境和生活习惯等因素的影响，呼吸系统疾病的患病率和死亡率居高不下。其中慢性阻塞性肺疾病因为早期症状不明显，所以是人们生命安全的潜在的威胁之一，因此对慢性阻塞性肺疾病进行早期筛查和预防尤为重要。本文通过对肺音信号进行去噪提升肺音信号的质量，提取并融合特征参数，最后通过深度学习对肺音信号进行分类。本文主要工作总结如下：

(1) 对采集的肺音信号进行降采样、去噪等预处理，得到相对纯净的肺音信号，便于更进一步地提取肺音信号特征。本文通过 IVMD 算法对肺音信号进行分解，对原始肺音信号先进行 VMD 分解，对含有噪声的固有模态分量继续进行 EMD 分解，去除噪声，然后根据相关系数的大小重构肺音信号，最后得到保留原始信号较多信息的去除噪声的纯净肺音信号。

(2) 提取肺音信号的线性预测倒谱系数和梅尔频率倒谱系数，并进行融合。利用肺音信号的本身特性，提取其 LPCC 和 MFCC，再将两种特征参数进行融合，得到包含更多肺音信号的信息的融合特征 LPCC-MFCC，为后续的分类奠定基础。

(3) 利用 CNN-Transformer 对肺音信号进行分类。首先采用结构适用性强的卷积神经网络作为主体框架进行分类任务，然后采用 CNN-SVM 网络模型，利用 SVM 替换 Softmax 函数泛化性，再利用 Transformer 提升网络的分类性能。实验结果表明，基于 CNN-Transformer 的肺音信号分类方法准确率为 95.70%，灵敏度为 94.62%、精确度为 93.87%、F1-score 为 93.88%，比 CNN-SVM 分别提升了 7.32%、2.49%、5.22%、1.51%。准确率更高，迭代速度更快。

### 6.2 展望

本文以“互联网+医疗健康”为研究基础，从肺音信号的微观特征出发，利用深度学习的方法构建了肺音信号分类预测模型，取得了部分成果，但是在实验过程中仍存在不足，后续研究有待进一步完善。

(1) 在数据集方面，本文采用的肺音信号均为单通道信号，且未对全部数据进行分析。大样本量和多通道的肺音信号有利于捕捉更多肺音信号的细节特征。因此，后期可以搭建多通道的肺音信号采集设备，建立专业数据库，分析研究更多的数据。

(2) 在信号的预处理和特征提取方面，本文采用 EMD、VMD 等算法进行去噪，去噪效果不是最理想的，且只提取了两种特征参数融合。现在已经有更多适用于非平稳非线性信号的处理方法，

可以根据肺音信号的特性进行更精确地去噪，因此后期可以采用参数化时频分析方法的 Chirplet 变换，建立更合适的数学模型逼近信号，从而提高模型精度，避免因迭代导致误差增大。同时提取更丰富的肺音信号特征进行评估并融合特征，精简特征参数，减少后期的模型计算任务。

(3)在肺音信号分类方面。本文以卷积神经网络为框架进行改进，但目前已经有 LSTM 等更新颖且适合时间序列的网络框架，后期可以将模型轻量化部署到嵌入式设备中应用于实际。本文进行健康人和 COPD 患者的区分，后续可以研究肺音信号多分类问题，提出适用面更广泛的分析系统，结合该方面的专家，为人们的健康问题提出更可行的肺音信号筛查系统。

## 参考文献

- [1] Li X, Cao X, Guo M, et al. Trends and risk factors of mortality and disability adjusted life years for chronic respiratory diseases from 1990 to 2017: systematic analysis for the Global Burden of Disease Study 2017[J]. *BMJ*, 2020(368): 1-10.
- [2] World cancer report: cancer research for cancer prevention[M]. IARC Press, 2020.
- [3] Sword D, Thomas K, Wise H, et al. A novel and cost-effective method for evaluating cardiopulmonary auscultation skills in student physical therapists[J]. *Journal of Allied Health*, 2017, 46(1): 9-13.
- [4] Jacome C, Marques A. Computerized Respiratory Sounds in Patients with COPD: A Systematic Review[J]. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, 2015, 12(1): 104-112.
- [5] Mondal A, Bhattacharya P, Saha G. Detection of Lungs Status Using Morphological Complexities of Respiratory Sounds[J]. *The Scientific World Journal*, 2014: 182938-182938.
- [6] 石陆魁, 刘文浩, 李站茹. 基于 LDA 和小波分解的肺音特征提取方法[J]. *计算机工程与应用*, 2017, 53(22):116-120+149.
- [7] 姚小静, 王洪, 刘盛雄. 基于遗传 BP 神经网络的肺音分类识别算法研究[J].*航天医学与医学工程*, 2016, 29(1): 45-51.
- [8] Pouyani M, Vali M, Ghasemi M. Lung sound signal denoising using discrete wavelet transform and artificial neural network[J]. *Biomedical Signal Processing and Control*, 2022, 72: 1-12.
- [9] Tripathy R, Dash S, Rath A, et al. Automated Detection of Pulmonary Diseases From Lung Sound Signals Using Fixed-Boundary-Based Empirical Wavelet Transform[J]. *IEEE Sensors Letters*, 2022, 6(5): 1-4.
- [10] 郭鹿鸣. 基于盲源分离的心肺音信号分离方法研究与应用[D]. 广东工业大学, 2016.
- [11] 朱俊霖, 王海平, 杨祖元.带标签约束的心肺音分离方法[J].*计算机工程*, 2019, 45(12): 314-320.
- [12] 雷志彬. 基于回归迁移非负矩阵分解和密集连接网络的心肺音分离[D]. 广东工业大学, 2019.
- [13] Muñoz-Montoro A, Suarez-Dou D, Cortina R, et al. Parallel source separation system for heart and lung sounds[J]. *The Journal of Supercomputing*, 2021, 77(8): 8135-8150.
- [14] Grooby E, He J, Fattahi D, et al. A new non-negative matrix co-factorisation approach for noisy neonatal chest sound separation[C]//2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Mexico, 2021, 5668-5673.
- [15] Aziz S, Khan M U, Shakeel M, et al. An Automated System towards Diagnosis of Pneumonia using Pulmonary Auscultations. MACS 2019–13th International Conference on Mathematics, Actuarial Science[J]. *Computer Science and Statistics, Proceedings*. 2019: 1-7.
- [16] Khan S, Pachori R. Automated classification of lung sound signals based on empirical mode decomposition[J]. *Expert Systems with Applications*, 2021, 184: 115456.
- [17] 高鸣蕾. 基于 EMD 的肺音信号去噪及特征提取方法研究[D]. 江苏大学, 2019.
- [18] 刘毅. 基于 EMD 技术的肺音信号处理关键技术研究[D]. 江苏大学, 2019.
- [19] 姚小静, 王洪, 李燕, 等. 肺音信号分析及其识别方法的研究进展[J]. *重庆理工大学学报*, 2013(012): 027.
- [20] 崔星星, 苏智剑. 一种新呼吸音信号特征提取方法与应用[J]. *中国医学物理学杂志*, 2018, 35(02): 214-218.

- [21] 黄晓旗, 闫苗苗, 王雷, 等. 基于呼气相不同阈值分析慢性阻塞性肺疾病严重程度与肺功能的相关性[J]. 中国医学影像技术, 2020, 36(03): 350-355.
- [22] 陈仕锋, 黄敏於, 彭显如, 等. 肺音可以作为首诊慢阻肺严重程度的判断指标[J]. 南方医科大学学报, 2020, 40(02): 177-182.
- [23] Infante C, Chamberlain D, Kodgule R, et al. Classification of voluntary coughs applied to the screening of respiratory disease[C]//2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Jeju, Korea, 2017: 1413-1416.
- [24] Rahman M, Nemati E, Rahman M, et al. Efficient online cough detection with a minimal feature set using smartphones for automated assessment of pulmonary patients[C]//Proceedings of the Ninth International Conference on Ambient Computing, Applications, Services and Technologies (AMBIENT 2019), Porto, Portugal. 2019: 22-26.
- [25] Abougabal M, Moussa N. A novel technique for validating diagnosed respiratory noises in infants and children[J]. Alexandria Engineering Journal, 2018, 57(4): 3033-3041.
- [26] Islam M, Bandyopadhyaya I, Bhattacharyya P, et al. Multichannel lung sound analysis for asthma detection[J]. Computer Methods and Programs in Biomedicine, 2018, 159: 111-123.
- [27] Göğüş F Z, Karlık B, Harman G. Identification of pulmonary disorders by using different spectral analysis methods[J]. International Journal of Computational Intelligence Systems, 2016, 9(4): 595-611.
- [28] Naqvi S, Choudhry M. An automated system for classification of chronic obstructive pulmonary disease and pneumonia patients using lung sound analysis[J]. Sensors, 2020, 20(22): 6512.
- [29] Porieva H, Ivanko K, Semkiv C, et al. Investigation of Lung Sounds Features for Detection of Bronchitis and COPD Using Machine Learning Methods[J]. Visnyk NTUU KPI Seria-Radiotekhnika Radioaparatobuduvannia, 2021 (84): 78-87.
- [30] Zhao X, Shao Y, Mai J, et al. Respiratory Sound Classification Based on BiGRU-Attention Network with XGBoost[C]//2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Seoul, Korea, 2020: 915-920.
- [31] 常峥, 罗萍, 杨波, 等. 基于 HHT-MFCC 和短时能量的慢性阻塞性肺病患者呼吸声识别[J]. 计算机应用, 2021, 41(2): 598-603.
- [32] Shanthakumari G, Priya E. Spectrogram-based detection of crackles from lung sounds[C]//2022 International Conference on Communication, Computing and Internet of Things (IC3IoT). Chennai, India, 2022: 1-6.
- [33] Haider N, Behera A. Computerized lung sound based classification of asthma and chronic obstructive pulmonary disease (COPD)[J]. Biocybernetics and Biomedical Engineering, 2022, 42(1): 42-59.
- [34] Rizal A, Puspitasari A. Lung sound classification using wavelet transform and entropy to detect lung abnormality[J]. Serbian Journal of Electrical Engineering, 2022, 19(1): 79-98.
- [35] Lozano M, Fiz J, Jané R. Automatic differentiation of normal and continuous adventitious respiratory sounds using ensemble empirical mode decomposition and instantaneous frequency[J]. IEEE journal of biomedical and health informatics, 2015, 20(2): 486-497.
- [36] Ulukaya S, Serbes G, Sen I, et al. A lung sound classification system based on the rational dilation wavelet transform[C]//2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Orlando, FL, USA, 2016: 3745-3748.
- [37] Fraiwan L, Hassanin O, Fraiwan M, et al. Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers[J]. Biocybernetics and Biomedical

- Engineering, 2021, 41(1): 1-14.
- [38] Lang R, Lu R, Zhao C, et al. Graph-based semi-supervised one class support vector machine for detecting abnormal lung sounds[J]. Applied Mathematics and Computation, 2020, 364: 124487.
- [39] Rizal A, Hidayat R, Nugroho H. Lung sound classification using Hjorth descriptor measurement on wavelet sub-bands[J]. Journal of Information Processing Systems, 2019, 15(5): 1068-1081.
- [40] Sengupta N, Sahidullah M, Saha G. Lung sound classification using cepstral-based statistical features[J]. Computers in biology and medicine, 2016, 75: 118-129.
- [41] Ma Y, Xu X, Li Y. LungRN+ NL: An Improved Adventitious Lung Sound Classification Using Non-Local Block ResNet Neural Network with Mixup Data Augmentation[J] Proc. Interspeech , 2020: 2902-2906.
- [42] Demir F, Sengur A, Bajaj V. Convolutional neural networks based efficient approach for classification of lung diseases[J]. Health Information Science and Systems, 2020, 8(1): 1-8.
- [43] Perna D. Convolutional neural networks learning from respiratory data[C]//2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Madrid, Spain, 2018: 2109-2113.
- [44] 包善书, 车波, 邓林红. 基于双源域迁移学习的肺音信号识别研究[J]. 计算机工程, 2023: 1-10.
- [45] Pham L, Phan H, Palaniappan R, et al. CNN-MoE based framework for classification of respiratory anomalies and lung disease detection[J]. IEEE Journal of Biomedical and Health Informatics, 2021, 25(8): 2938-2947.
- [46] Chen H, Yuan X, Pei Z, et al. Triple-classification of respiratory sounds using optimized s-transform and deep residual networks[J]. IEEE Access, 2019, 7: 32845-32852.
- [47] Nguyen T, Pernkopf F. Lung sound classification using co-tuning and stochastic normalization[J]. IEEE Transactions on Biomedical Engineering, 2022, 69(9): 2872-2882.
- [48] Jayalakshmy S, Sudha G. GTCC-based BiLSTM deep-learning framework for respiratory sound classification using empirical mode decomposition[J]. Neural Computing and Applications, 2021, 33(24): 17029-17040.
- [49] Neto J, Arrais N, Vinuto T, et al. Convolution-Vision Transformer for Automatic Lung Sound Classification[C]//2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). Natal, Brazil, 2022: 97-102.
- [50] Huang Y, Song I. PhysioVec: A Multi-stage Deep-Learning Framework for Searching Online Health Information with Breath Sound[C]//2022 IEEE 5th International Conference on Big Data and Artificial Intelligence (BDAI). Fuzhou, China, 2022: 26-33.
- [51] Barbouchi K, El Hamdi D, Elouedi I, et al. A transformer - based deep neural network for detection and classification of lung cancer via PET/CT images[J]. International Journal of Imaging Systems and Technology, 2023: 1-13.
- [52] Sun W, Pang Y, Zhang G. CCT: Lightweight compact convolutional transformer for lung disease CT image classification[J]. Frontiers in Physiology, 2022, 13: 1-13.
- [53] 陈仕锋. 肺音可以作为首诊慢阻肺严重程度的判断指标[D]. 南方医科大学, 2021.
- [54] 郑明杰. 基于希尔伯特—黄变换的肺音信号识别方法研究[D]. 江苏大学, 2016.
- [55] Kim Y, Hyon Y, Lee S, et al. The coming era of a new auscultation system for analyzing respiratory sounds[J]. BMC Pulmonary Medicine, 2022, 22(1): 1-11.
- [56] Rosa M, Seidel H, Paim G, et al. An energy-efficient haar wavelet transform architecture for respiratory signal processing[J]. IEEE Transactions on Circuits and Systems II: Express Briefs,

- 2020, 68(2): 597-601.
- [57] Chatziparasidis G, Priftis K, Bush A. Wheezing as a Respiratory Sound[J]. Breath Sounds: From Basic Science to Clinical Practice, 2018: 207-223.
- [58] Fraiwan M, Fraiwan L, Alkhodari M, et al. Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory[J]. Journal of Ambient Intelligence and Humanized Computing, 2021: 1-13.
- [59] Sarkar M, Bhardwaz R, Madabhavi I, et al. Physical signs in patients with chronic obstructive pulmonary disease[J]. Lung India: Official Organ of Indian Chest Society, 2019, 36(1): 38-47.
- [60] Goodman D, Crocker M, Pervaiz F, et al. Challenges in the diagnosis of paediatric pneumonia in intervention field trials: recommendations from a pneumonia field trial working group[J]. The Lancet Respiratory Medicine, 2019, 7(12): 1068-1083.
- [61] 韦海成, 冯海青, 塔娜, 等. 基于 VMD 分解的呼吸音频谱特征在 COPD 识别中的应用[J]. 现代电子技术, 2021, 44(19): 61-65.
- [62] Dragomiretskiy K, Zosso D. Variational Mode Decomposition[J]. IEEE Transactions on Signal Processing, 2014, 62(3): 531-544.
- [63] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017: 5998–6008.
- [64] Moore B C J. Effects of hearing loss and age on the binaural processing of temporal envelope and temporal fine structure information[J]. Hearing research, 2021, 402: 1-13.
- [65] Rocha B, Filos D, Mendes L, et al. A respiratory sound database for the development of automated classification[C]//International Conference on Biomedical and Health Informatics. Springer, Singapore, 2017: 33-37.
- [66] Demir F, Ismael A, Sengur A. Classification of lung sounds with CNN model using parallel pooling structure[J]. IEEE Access, 2020, 8: 105376-105383.
- [67] Petmezas G, Cheimariotis G, Stefanopoulos L, et al. Automated Lung Sound Classification Using a Hybrid CNN-LSTM Network and Focal Loss Function[J]. Sensors, 2022, 22(3): 1232.