



本科生毕业设计[论文]

基于半监督学习的心音异常检测和分类

院 系 _____

专业班级 _____

姓 名 _____

学 号 _____

指导教师 _____

_____ 年 月 日

学位论文原创性声明

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的
研究成果。除了文中特别加以标注引用的内容外，本论文不包括任何其他个人或集
体已经发表或撰写的成果作品。本人完全意识到本声明的法律后果由本人承担。

作者签名： 年 月 日

学位论文版权使用授权书

本学位论文作者完全了解学校有关保障、使用学位论文的规定，同意学校保留
并向有关学位论文管理部门或机构送交论文的复印件和电子版，允许论文被查阅
和借阅。本人授权省级优秀学士论文评选机构将本学位论文的全部或部分内容编
入有关数据进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论
文。

本学位论文属于 1、保密，在 年解密后适用本授权书
2、不保密。

(请在以上相应方框内打“√”)

作者签名： 年 月 日
导师签名： 年 月 日

摘 要

当前深度学习已广泛用于心音的异常检测。这些深度学习方法往往需要大量的带标签数据，然而心音的标注工作需要大量的医生资源且费时费力，对带标签数据的依赖性在一定程度上限制了当前方法的性能。而无标签的心音数据是易获取的，如果能够利用无标签数据减小对带标签数据的依赖，将极大提高基于深度学习的心音异常检测方法的临床实用性。因此，本文将研究半监督方法在心音异常识别中的应用，提出一种基于半监督学习、能够有效减小带标签数据依赖的心音异常检测方法。主要包括以下内容：

（1）本文根据心音信号特点设计了适用于心音信号的预处理方法，包括心音信号采集过程中产生的噪声的滤除、受试者之间信号强度差异的消除等，为深度学习应用于心音信号异常检测和分类奠定基础。

（2）本文结合心音异常检测的任务特点，提出基于残差网络的 P-ResNet 网络框架，并在 George B.Moody PhysioNet 2022 挑战赛数据集上评估模型性能。结果表明，本文设计的 P-ResNet 网络 AUC 值达到 0.8950，实现先进的性能。

（3）本文以 P-ResNet 为网络骨干，将传统平均教师模型与更新锐化方法结合，设计适用于心音信号的基于混合教师模型的半监督学习方法。本文设计的混合教师方法仅需 25%带标签数据便可实现与使用全部带标签数据的全监督方法相近的性能，表明该方法可充分利用无标签数据，极大减小对带标签数据的依赖。

综上所述，本文首先根据心音信号特点设计了合适的预处理方法，为心音异常检测和分类奠定基础；然后，本文提出了 P-ResNet 网络，能够有效检测异常心音信号；在此基础上，本文进一步以 P-ResNet 为网络骨干开发了基于混合教师模型的半监督学习方法，可以极大减小对带标签数据的依赖，提高了深度学习在心音异常检测上的临床实用性。

关键词：心音异常检测；深度学习；P-ResNet；半监督学习；混合教师

Abstract

Phonocardiogram (PCG) is a significant tool for detecting cardiovascular disease. However, identifying abnormal heart sounds remains a significant challenge due to the high frequency and instability of the audio signal, relying heavily on the subjective expertise of physicians. In developing countries, experienced physicians are scarce. As a result, there is a great need for an objective and accurate method for the automatic detection of abnormal heart sounds.

Recently, Deep learning has been widely used for abnormal heart sounds detection which requires quantities of labeled data. Besides, labeling heart sounds is time-consuming and laborious for physicians. And the reliance of labeled data limits the performance of current methods to a certain extent. In the contrast, unlabeled data is more easily accessible. So reducing reliance on labeled data can greatly improve the clinical utility. This paper investigates the application of semi-supervised methods for heart sound abnormality detection and proposes a semi-supervised learning-based method that effectively reduces dependence on labeled data including following contents.

(1) This paper presents a preprocessing method tailored to heart sound signals, including filtering noise generated during signal acquisition and eliminating signal intensity differences between patients. This lays the foundation for applying deep learning to heart sound signal abnormality detection and classification.

(2) Additionally, according to the task characteristics of heart sound abnormality detection, a P-ResNet network framework based on residual networks is proposed. Model performance is evaluated on the George B. Moody PhysioNet 2022 Challenge dataset, with results demonstrating advanced performance with an AUC value of 0.8950.

(3) Using P-ResNet as the cornerstone of the network, this paper proposes a semi-supervised learning method based on the mixed teacher model. This approach combines

the traditional mean teacher model with updating and sharpening methods, achieving similar performance to the fully supervised method using only a quarter of the labeled data. This shows that the method can make full use of unlabeled data and greatly reduce reliance on labeled data, improving the clinical utility of deep learning for abnormal heart sound detection.

Overall, this paper presents a suitable preprocessing method, a P-ResNet network, and a semi-supervised learning method based on the mixed teacher model, which can effectively reduce reliance on labeled data and improve the performance of deep learning methods for abnormal heart sound detection.

Key Words: Heart sound abnormality detection; Deep learning; P-ResNet; Semi-supervised learning; Mixed teacher

目 录

摘 要	I
Abstract.....	II
1 绪论	1
1.1 背景及意义.....	1
1.2 国内外研究现状.....	2
1.3 研究内容及章节安排.....	4
2 心音信号预处理和深度学习网络基础	6
2.1 数据集.....	6
2.2 预处理.....	6
2.3 卷积神经网络结构.....	8
2.2.1 卷积层.....	8
2.2.2 批量归一化.....	9
2.2.3 激励层.....	11
2.2.4 池化层.....	12
2.2.5 全连接层.....	13
2.4 本章小结.....	14
3 基于全监督学习的心音异常检测和分类	15
3.1 全监督网络.....	15
3.1.1 残差模块.....	15
3.1.2 P-ResNet 网络结构	16
3.2 训练和测试.....	18
3.2.1 数据划分.....	18
3.2.2 模型训练.....	19
3.2.3 模型测试.....	19

3.2.4 常见评价指标.....	20
3.3 实验结果.....	22
3.4 本章小结.....	22
4 基于半监督学习的心音异常检测和分类	23
4.1 半监督方法.....	23
4.2 训练和测试过程.....	25
4.3 实验结果.....	26
4.3.1 半监督学习方法与全监督学习方法性能对比.....	26
4.3.2 关键参数与模块分析.....	28
4.4 本章小结.....	30
5 总结和展望	31
5.1 总结.....	31
5.2 展望.....	32
致谢	33
参考文献	34

1 绪论

1.1 背景及意义

据世界卫生组织统计^[1]，2019 年全球估计有 1790 万人死于心血管疾病，占全球死亡人数的 32%，超过四分之三的心血管疾病死亡发生在中低收入国家，预计到 2030 年这一数字将超过 2300 万，心血管疾病已经成为威胁人类健康的“头号杀手”。我国也面临着同样的情况，我国心血管疾病死亡率呈现逐年上升趋势，高于肿瘤及其他疾病；据国家心血管病中心推算，我国心血管疾病现有患者人数约 3.3 亿^[2]。心血管疾病的早期检查和诊断对于患者疾病治疗、提升全社会整体健康水平具有十分重要的意义。

现有心血管疾病检查方式主要有：心脏听诊（记录为心音图 Phonocardiogram, PCG）、心电图（Electrocardiogram, ECG）、心脏彩超、心脏血管造影、心脏电子计算机断层扫描(CT)等。其中，心音信号是在人体生理活动过程中自发产生的，无需任何外界干预，与心脏彩超、心脏血管造影、心脏 CT 等借助外部能量源的检查方式相比，具有无创、便捷、高效且经济的特点，更适合用于心血管疾病的早期检查和诊断，同时也可用于心血管疾病的预防以及病情的长期监测等。

心音主要是由心脏瓣膜关闭而产生的，在一个心动周期中主要包括两个基础心音。在收缩期开始时，房室瓣（三尖瓣和二尖瓣）关闭产生第一心音（S1），在收缩期结束的时候，主动脉瓣和肺动脉瓣关闭产生第二心音（S2）^[3]。

心音信号对心脏产生的声音进行了高保真的记录，反映了心血管系统运动状态，包含着心血管疾病的早期病理信息。心音信号擅长检测变力性心血管疾病，异常心音往往与疾病有关^[4]，如心脏瓣膜疾病^[5,6]、心内间隔缺损、动脉导管未闭^[7]等，在心血管疾病的早期检测中扮演着重要的角色。通过心音信号，医生可以了解心脏的健康情况，并发现潜在的疾病。但是心音信号本身具有高频，不稳定等特点，通过专家人工解读心音信号费时费力，而且非常依赖专家们的临床经验，因此，开发一种客观、高可靠性、高精度的心音信号自动检测和分类方法非常重

要。

随着计算机技术的发展,深度学习已经广泛用于心音信号异常检测,并取得了良好的性能。但是这些方法通常需要大量的带标签数据进行训练,而带标签数据的获取需要占用大量的医生资源,费时耗力,这在一定程度上限制了当前方法的实用性。相比之下,无标签数据获取成本较低,如果能够通过有效地利用无标签数据而减少对带标签数据的需求,将有助于提高心音异常自动检测方法的实用性。因此,开发一种对带标签数据依赖性低且适用于心音异常检测的半监督学习方法十分重要。

1.2 国内外研究现状

近年来,已经提出了许多用于心音信号分类的自动化方法,这些方法在临床检测病理学中展现出巨大的潜力。2016 年和 2022 年举办的 PhysioNet challenge 旨于心音信号的准确分类^[8,9],如心音信号中杂音的检测等。

在过去的研究中,为了实现心音信号的准确分类,经典的机器学习(Machine Learning)方法已被广泛使用,Whitaker^[10]等人使用稀疏编码方法提取 PCG 信号的频域特征,同时结合 PCG 信号的时域特征,对支持向量机(SVM)进行训练,而 Cornely^[11]等人则通过小波散射变换提取 PCG 信号的频域特征训练支持向量机,实现心音信号的分类。然而,经典机器学习通常在小规模数据上表现很好,在大规模数据集上的性能往往具有局限性^[12]。

近年来,深度学习(Deep learning)方法在心音信号分类上取得了更加强大的性能,体现了其比传统机器学习更强大的心音信号分析能力^[13]。深度学习模型通常使用原始音频信号或者从心音信号中提取到的时频表征作为输入^[14,15],同时深度学习模型更复杂的结构也使其能够从大规模数据集中学习以提高性能。以下为目前致力于心音信号分类的主流深度学习方法。

(1) 采用时频表征的深度学习方法

由于从原始心音信号中手工设计和提取特征具有一定的挑战性,因此二维的时频表征常被作为二维卷积神经网络(CNN)的输入^[16-20]。Balamurugan^[21]等人

通过短时傅里叶变换（STFT）从心音信号中提取时频谱送入 ResNet 网络中检测异常心音信号。

（2）用其他特征的深度学习方法

除了上述采用时频表征的深度学习方法之外，还有诸多采用其他特征的深度学习方法。

① 时域特征

与传统机器学习中提取的特征类似，许多深度学习方法提取一维时域特征作为模型的输入。例如，从每个 75ms 的心音信号片段中提取多个统计特征值（如均值，中值，方差等）并送入双向 LSTM（BiLSTM）网络进行训练^[22]。

② 基于频域的一维特征

一维 CNN 或前馈 DNN 模型均可用于处理一维特征。Ranipa^[20]等人将频率特征和梅尔域特征送入一维 CNN 中，然后连结多个 CNN 得到最终预测结果。

③ 基于频域的二维特征

Qian^[23]等人利用小波变换从一组声学片段中计算出小波能量特征，并使用 GRU-RNNs 作为分类器。Dong^[24]等人从心音片段中提取 log Mel 特征，并使用 LSTM-RNNs 和 GRU-RNNs 实现分类。Zhang^[25]等人则从频谱图中提取由平均幅度差函数计算得来的时间周期特征，并使用 LSTM-RNNs 来探索特征内的相关性。

（3）端到端的深度学习方法

近年来，由于时频表征等仍然需要手动设计和选择，可以直接从原始心音信号中学习的端到端网络已经逐渐成为一种新颖有效的方法。目前基于原始心音信号，研究人员已经提出多种 1D CNN 框架来处理心音信号分类任务^[26-29]。Liu^[30]等人，提出了一种在心音分类任务中体现出高灵敏度的时间卷积网络（TCN），而 S. L. Oh^[31]等人则开发了一种由残差模块组成的一维 CNN 模型用于心音信号分类。Shuvo^[32]等人提出了 CardioXNet 模型，在表征学习阶段，三个并行的 1D CNN 通道来从心音信号中提取时不变特征，在序列残差学习阶段，使用 BiLSTM 来学习序列表征。Li^[33]等人使用 1D CNN 来提取频域特征，而时域特征由 GRU-

RNN 提取，并进一步使用自注意力机制来融合两类特征来进行最终分类。

上述全监督方法在心音分类的问题上已经取得了良好的性能，但是由于标注异常心音信号费时费力，开发用于心音信号分类的半监督学习方法仍具有很高的价值

半监督学习方法结合全监督学习方法和无监督学习方法，通过使用无标签数据来显著减少对带标签数据的需求^[34]。随着深度学习的发展，已经诞生了许多半监督学习方法^[35]。这些方法大体可分为五类：一致性正则化方法、生成方法、基于图的方法、伪标签方法和混合方法^[36,37]。其中时间集成^[38]和平均教师方法^[39]是一致性正则化方法的代表。半监督 GAN 和半监督 VAE 属于生成方法。基于图的神经网络的方法，自训练模型^[40]和插值一致性训练^[41]则属于基于图的方法，伪标签方法和混合方法。目前，尚未在临床心音信号分类任务中开发有效的半监督学习方法。

1.3 研究内容及章节安排

本文主要针对基于深度学习的心音异常检测展开研究。首先针对心音信号特点设计了恰当的心音信号预处理方法，为后续研究奠定基础，接着，在经过广泛调研后，本文设计了 P-ResNet 模型作为本次研究的核心模型，在此基础上，开发了基于混合教师模型的半监督学习方法，使得模型在半监督学习方法下，仅使用四分之一带标签数据，即可与全监督方法下使用全部带标签数据的模型性能相近，极大地减轻了模型训练对带标签数据的依赖。最后，对本文工作进行了总结和展望。本文一共分为五个章节，章节安排如下：

第一章：绪论。本章简要说明了心音信号异常检测和分类的研究意义及其研究背景，同时总结了心音信号异常检测和分类的国内外研究现状。

第二章：预处理和深度学习网络基础。本章首先介绍了本文采用的数据集和数据预处理部分的工作，之后介绍了本文所采用的深度学习网络架构的基本知识。

第三章：全监督学习的心音异常检测和分类模型。本章介绍了本研究采用的核心模型框架：P-ResNet，并简要阐述了其作用机制和实现原理。接着，通过全

监督学习的方法对模型进行充分训练，评估并展示了模型的性能。

第四章：半监督学习的心音异常检测和分类模型。本章首先阐述了本文采用的半监督学习方法：混合教师方法。接着，本章通过调整损失权重、设计加噪方式以及更新锐化等系列改进措施，进一步提升模型性能，使模型仅需四分之一带标签数据便可实现与使用全部带标签数据的全监督方法相近的性能。

第五章：总结和展望。本章总结了本研究在数据预处理，全监督学习和半监督学习这三项任务上的工作，同时针对本文的局限性，对下一步工作方向进行了展望，有助于开展下一步的工作。

2 心音信号预处理和深度学习网络基础

2.1 数据集

本文所采用的数据集来自于 George B.Moody PhysioNet 2022 挑战赛^[8]。数据来源于多个听诊位置，包括肺动脉瓣（PV）、主动脉瓣（AV）、二尖瓣（MV）以及三尖瓣（TV），记录的数量、位置和持续时间因患者而异。

在本次实验中，我们共使用来自 874 名患者的 2982 条心音信号记录，其中 731 条来自肺动脉瓣（PV）听诊位置，748 条来自主动脉瓣（AV）听诊位置，803 条来自二尖瓣（MV）听诊位置，以及 700 条来自三尖瓣（TV）听诊位置。这些数据记录为我们的研究提供了丰富的样本信息，能够全面地反映心脏疾病的特征和规律。

我们使用的心音数据集主要分为两类，分别是“Absent”和“Present”。其中，“Absent”类别表示病人心音中不存在杂音，我们将其标签值设置为“0”；“Present”类别则表示病人心音中存在杂音，我们将其标签值设置为“1”。在数据集中，每个患者可能存在来自一个或多个不同采样位点的心音信号。为了保证数据的统一性，我们根据患者的标签情况，将这些心音信号赋予相应的标签值。

2.2 预处理

在信号处理过程中，常使用滤波器来去除信号中的噪声等干扰，可以使信号中特定频率成分被消除或减少。常用的滤波器有：巴特沃兹滤波器，切比雪夫滤波器，椭圆滤波器，贝塞尔滤波器等。巴特沃兹在通频带内的频率响应曲线最大限度平坦，没有纹波，而在阻频带则逐渐下降为零；切比雪夫滤波器在通带或阻带上频率响应幅度等波纹波动；椭圆滤波器又称考尔滤波器，在通带或阻带等波纹波动，相比其他类型的滤波器，在阶数相同的条件下有着最小的通带和阻带波动；贝塞尔滤波器具有最大平坦的群延迟（线性相位响应），在通频带内，各种频率的信号经滤波器后产生不同相移，相移与频率成线性关系，使波形失真最小。

在本文中，读取信号时，我们首先对信号采取了幅度归一化的处理，之后根据心音信号特点，我们采用了五阶巴特沃兹滤带通滤波器对心音信号进行滤波，它可以通过削弱信号中的高频和低频成分，使信号变得平滑。作为一种线性滤波器，它具有平滑的频率响应和较快的滚降特性。在其通频带内，其频率响应曲线是最大限度平坦的，而在其阻频带内，其频率响应则逐渐下降至零。

由于心音有效信息的频率范围主要在 400Hz 以下，同时为了滤除基线漂移等低频干扰，本文将带通滤波器截止频率分别设置为 25Hz 和 400Hz。这样的设置能够有效地滤除信号中的高频和低频噪声，减少外部干扰。除此之外，我们进一步对心音信号进行了分帧处理，对于一段完整的心音信号，我们采用窗口长度 10000，滑动距离 5000 的滑动窗，将心音信号处理为等长的心音信号片段。

经过以上处理，我们最终得到了归一化之后等长的 50132 条心音片段用于深度学习训练，其中共有阳性片段 9917 条，阴性片段 40215 条，在接下来的研究中，我们再把处理好的心音信号片段送入分类器中，实现心音信号的异常检测和分类，整体流程见图 2-1。该处理方法能够有效地滤除信号中的干扰和噪声，提高了信号的可靠性和稳定性。

在数据预处理的过程中，我们对一段完整的心音数据进行了幅度归一化、滤波和分帧的处理，将其划分为多个归一化的等长心音片段。对于这些片段，我们同样赋予了和心音记录相同的标签值，使得心音片段的标签值与其所属的心音记录的标签值保持一致，从而确保训练模型的准确性和可靠性。

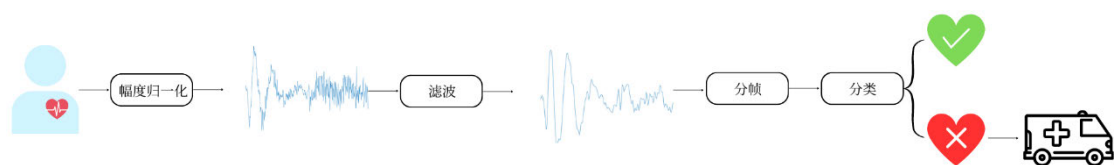


图 2-1 心音信号分类流程

2.3 卷积神经网络结构

卷积神经网络作为深度学习领域最具代表性的神经网络之一，已经取得了辉煌的成就，被广泛运用于人脸识别、自动驾驶汽车、自助超市和智能医疗等多项重要研究之中，而在本文中，我们同样使用卷积神经网络用于心音信号的检测和分类。

现代卷积神经网络的发展主要始于 2012 年，Alex^[42]等人使用 CNN 在 ImageNet Large Scale Visual Recognition Challenge (LSVRC) 中取得了最佳成绩，引起了广泛的关注，极大地推进了现代卷积神经网络的发展。

卷积神经网络作为一种前馈神经网络，不同于传统的手工特征提取方法，卷积神经网络架构可以自动从数据中提取特征，实现了强大的性能，接下来将详细介绍在本实验中具体使用的网络结构。

2.2.1 卷积层

卷积层是深度学习中最核心的层之一，主要用于对输入数据进行特征提取等操作。卷积层的内部通常包含多个卷积核，通过卷积核对输入特征图进行滤波操作，从而得到一个新的特征图作为输出。而卷积核通常由一系列可训练的权重组成。卷积区域的大小取决于卷积核的大小，也常常被称为感受野。卷积层的核心思想是局部连接和权值共享。具体地，卷积核在输入特征图上进行滑动操作，每次取出一个局部区域的值，然后通过卷积核中的权重计算得到输出特征图中对应位置的值。由于卷积核在输入特征图上的滑动操作，在一定程度上实现了对输入数据的平移不变性，同时通过权值共享可以大大减少模型的参数量，提高了模型的泛化能力和计算效率。

以二维卷积核为例，卷积运算过程如下

$$Z^{l+1} = (Z^l * g)(m, n) = \sum_i \sum_j Z^l(i, j) g(m - i, n - j) \quad (2-1)$$

$$L_{l+1} = \frac{L_l + 2p - f}{s_0} + 1 \quad (2-2)$$

其中 Z^l 和 Z^{l+1} 是第 l 层卷积的输入和输出，也称为特征图， g 为卷积核， L_{l+1} ，

L_l 分别为 Z^{l+1} 和 Z^l 的尺寸， f ， s_0 和 p 为卷积层参数，对应卷积核大小，卷积步长和填充层数。

卷积层参数包含卷积层大小，步长和填充，三者共同决定了卷积层输出特征图的尺寸，是卷积神经网络的超参数。卷积核越大，即感受野越大，可提取的输入特征越复杂。卷积步长定义了卷积核相邻两次扫过特征图时位置的距离，当步长为 1 时，会逐个扫过特征图的元素。填充作为深度学习中常用的一种技术，用于在输入数据的周围添加一些额外的数据，在卷积操作中，卷积核的大小通常小于输入数据的大小，因此，当卷积核移动到输入数据的边缘时，可能出现边缘像素没有被完全计算到的情况，这会导致输出数据的大小比输入数据小，丢失一部分信息。为了避免这种情况的发生，我们可以在输入数据的周围添加一些额外的数据，使得卷积核能够完全覆盖输入数据的所有像素点，从而保留更多的信息。填充有两种类型，一种是 Valid padding，另一种是 Same padding。Valid padding 指的是不添加额外的数据，直接对输入数据进行卷积操作，输出数据的大小会根据卷积核的大小和步长来进行计算。而 Same padding 则是在输入数据的周围添加额外的数据，使得输出数据的大小与输入数据的大小相同，通常会在卷积核的边缘添加一些 0 值像素点，称为“0 padding”。填充的作用不仅仅限于卷积层，同样也适用于池化层等操作。在池化操作中，池化核的大小通常也小于输入数据的大小，因此也可能出现像素点没有被完全计算到的情况。在这种情况下，使用填充可以保证池化核能够完全覆盖所有的像素点，提高模型的表现。

2.2.2 批量归一化

批量归一化（Batch Normalization）是深度学习中常用的一种技术，用于规范化神经网络中每层的输入，从而减少内部协变量位移（Internal Covariate Shift）现象的影响，提高模型的训练效率和稳定性。

在深度神经网络中，每一层的输入都会受到前一层输出的影响，导致每一层的输入分布发生变化。这种现象被称为内部协变量位移。内部协变量位移会使得模型训练变得困难，因为每一层的参数都需要重新适应输入分布的变化。Batch

Normalization 通过对每个 batch 的输入进行规范化，即对每个特征维度进行零均值化和方差归一化，使得每一层的输入分布更加稳定，从而减少内部协变量位移的影响。

池化层可以被添加到神经网络的卷积层或全连接层中。在卷积层中，Batch Normalization 通常被添加到卷积操作和激励函数之间，对卷积输出进行规范化。在全连接层中，Batch Normalization 被添加到线性变换和激励函数之间。Batch Normalization 的具体计算过程如下：

$$B = \{x_{1...m}\} \quad (2-3)$$

$$y_i = BN_{\gamma, \beta}(x_i) \quad (2-4)$$

$$\mu_B = \frac{1}{m} \sum_{i=1}^m x_i \quad (2-5)$$

$$\delta_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \quad (2-6)$$

$$\tilde{x}_i = \frac{x_i - \mu_B}{\sqrt{\delta_B^2 + \varepsilon}} \quad (2-7)$$

$$y_i = \gamma \tilde{x}_i + \beta \quad (2-8)$$

其中 B 表示每一个批次送入的数据， γ, β 为 batch normalization 层可以训练的参数， μ_B 和 δ_B^2 为 B 的均值和方差， B 中数据经过归一化之后得到 \tilde{x}_i ，最后将归一化上的数据通过可训练参数还原特征。

Batch Normalization 的作用可以从以下三个方面来解释：

(1) 减少内部协变量位移

由于每一层的输入分布发生变化，导致模型训练变得困难，Batch Normalization 通过对每个 batch 的输入进行规范化，使得每一层的输入分布更加稳定，从而减少内部协变量位移的影响，提高模型的训练效率和稳定性以及泛化能力。

(2) 减少对初始权重的依赖

Batch Normalization 可以减少模型对初始权重的依赖，从而提高模型鲁棒性，

并且可以提高模型的泛化能力。这是因为 Batch Normalization 对每个 batch 进行规范化，可以使得每个特征维度的均值和方差都在一个较小的范围内，从而减少特征之间的相关性，提高模型的泛化能力。

（3）改善模型的收敛速度和性能

Batch Normalization 可以使得每一层的输入分布更加稳定，从而减少模型训练过程中的梯度消失和梯度爆炸现象，改善模型的收敛速度和性能。此外，Batch Normalization 还可以增加模型的正则化效果，降低模型的过拟合风险。

2.2.3 激励层

激励层（Activation Layer）是神经网络中的一种常用层，其作用是在神经网络的每个神经元输出上引入非线性变换，以增加模型的表达能力和学习能力。

在神经网络中，每个神经元的输出是由输入加权和经过激励函数处理得到的。激励函数通常是非线性的，可以使得神经网络具有更强的表达能力和学习能力。如果没有激励函数，神经网络将只能进行线性变换，无法处理非线性问题。

激励层的作用可以通过以下三个方面来解释。

（1）引入非线性变换

神经网络中的每个神经元都是一个非线性单元，其输出是由输入加权和通过激励函数处理得到的。激励函数通常是非线性的，可以使得神经网络具有更强的表达能力和学习能力。

（2）解决线性可分问题

对于一些线性可分的问题，比如逻辑回归，线性回归等，神经网络只需要一层激励层就可以解决。激励函数将神经网络的输出限制在一定的范围内，从而可以处理线性可分问题。

（3）实现非线性映射

神经网络中的每个神经元都可以看作是一个非线性映射，将输入映射到输出空间中。不同的激励函数可以实现不同的非线性映射，从而增加神经网络的表达能力和学习能力。例如，ReLU 激励函数可以实现稀疏表示，提高模型的泛化能

力；Sigmoid 激励函数可以将输出限制在[0,1]之间，可以用于二分类问题等。

常见的激励函数包括 Sigmoid、ReLU、Tanh 等。其中，Sigmoid 函数具有平滑的 S 形曲线，可以将输出限制在[0,1]之间，常用于二分类问题；ReLU 函数具有简单的形式和快速的计算速度，可以实现稀疏表示，常用于卷积神经网络中；Tanh 函数具有类似于 Sigmoid 函数的形状，但输出范围在[-1,1]之间，常用于回归问题。

卷积层中包含激励函数来协助表达特征，在本研究中使用的激励函数是 ReLU 函数，表达式为

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (2-9)$$

2.2.4 池化层

卷积层进行特征提取后，输出的特征图会被传递至池化层进行特征选择和信息过滤。池化层（Pooling Layer）包含预设定的池化函数，其功能是将特征图中单个点的结果替换为其相邻区域的特征图统计量。其作用是对输入特征图进行下采样，减少特征图的尺寸和参数数量，并提取输入特征图的主要信息，增强模型的鲁棒性和泛化能力。

池化层的作用可以通过以下三个方面来解释：

（1）减少特征图的尺寸和参数数量

池化层可以通过对输入特征图进行下采样，减少特征图的尺寸和参数数量，从而加速模型训练，减少过拟合的风险。池化层通常采用固定的窗口大小和步幅，对每个窗口内的值进行池化操作，得到一个池化值，从而减少特征图的尺寸和参数数量。

（2）提取输入特征图的主要信息

池化层可以通过对输入特征图进行下采样和池化操作，提取输入特征图的主要信息，从而增强模型的鲁棒性和泛化能力。池化层通常采用最大池化（Max Pooling）或平均池化（Average Pooling）等操作，可以提取输入特征图中的主要特征。

(3) 增强模型的鲁棒性和泛化能力

池化层可以通过减少特征图的尺寸和参数数量,提取输入特征图的主要信息等操作,增强模型的鲁棒性和泛化能力。池化层可以使得模型对输入数据的微小变化具有较强的鲁棒性,从而提高模型的泛化能力。

在神经网络中,输入特征图通常具有较大的尺寸和参数数量,这会导致模型训练变得困难,容易导致过拟合。池化层可以通过对输入特征图进行下采样,减少特征图的尺寸和参数数量,从而加速模型训练,减少过拟合的风险,并提取输入特征图的主要信息,增强模型的鲁棒性和泛化能力。

池化层通过将一层神经元簇的输出组合到下一层中的单个神经元来减少数据的维度,流行的池化有两种:最大值和平均值。**Max pooling** 使用特征图中每个局部的最大值,而 **Average pooling** 取平均值。

本研究使用的是最大池化层 (**Max Pooling**), 其定义为在每个池化窗口内选择最大值作为池化值。最大池化层是一种常用的池化操作。在本研究中,使用 2×2 , 步长为 1 的最大池化层, 其中 X 表示最大池化层的输入, Y 表示最大池化层的输出, X 为 $m \times n$ 矩阵

$$Y = \text{maxpooling}(X) \quad (2-10)$$

$$y_{i,j} = \max(x_{i,j}, x_{i,j+1}, x_{i+1,j}, x_{i+1,j+1}) \quad (2-11)$$

$$i \leq m - 2 \quad j \leq n - 2 \quad (2-12)$$

2.2.5 全连接层

全连接层 (**Fully Connected Layer**) 位于卷积神经网络隐含层的最后部分, 并只向其它全连接层传递信号。特征图在全连接层中会失去空间拓扑结构, 被展开为向量, 卷积神经网络中的卷积层和池化层能够对输入数据进行特征提取, 全连接层的作用则是对提取的特征进行非线性组合以得到输出, 它可以将特征映射到一个抽象空间并将其分离, 即全连接层本身不被期望具有特征提取能力, 而是试图利用现有的高阶特征完成学习目标。全连接层通常采用向量乘法和偏置项的方式计算输出值, 其中向量乘法用于实现特征变换, 偏置项用于增加非线性性和平

移不变性。全连接层公式如下，其中 X 表示最大池化层的输入， Y 表示最大池化层的输出， W 为全连接层权重， b 表示全连接层偏置。

$$Y = WX + b \quad (2-13)$$

2.4 本章小结

本章我们首先介绍了采用的数据集的基本情况，接着，针对心音信号特点，详细介绍了提出的心音信号预处理方法，我们通过幅度归一化，滤波以及分针等方法将心音信号划分为适用于后续深度学习研究的心音片段。同时，还介绍了卷积神经网络的基本结构和框架，为下文中开发适用于心音信号分类的深度学习网络框架奠定基础。

3 基于全监督学习的心音异常检测和分类

3.1 全监督网络

深度学习网络中常常存在梯度消失和梯度爆炸等问题,使得深层次的网络难以训练。为此,He^[43]等人提出了残差模块(Residual Block)以及深度残差模型(ResNet),其通过引入残差块的方式,使得梯度能够直接传递,避免了深度网络中的梯度问题,使得模型能够在较深的网络结构下学习,从而提高模型性能。

3.1.1 残差模块

在深度神经网络中,随着网络层数的增加,梯度消失和梯度爆炸问题变得越来越严重。这是由于在反向传播过程中,梯度会被多次连乘,从而导致梯度值变得非常小或非常大,使得模型难以训练和优化。针对这一问题,在深度学习任务中常常需要引入残差模块。

残差模块的作用可以通过以下几个方面来解释:

(1) 解决梯度消失和梯度爆炸问题

残差模块通过引入跨层连接的方式,将前一层的输出直接添加到当前层的输入中,从而实现梯度的直接反向传播,解决梯度消失和梯度爆炸问题。这可以使模型更加容易训练和优化,并提高模型的鲁棒性和泛化能力。

(2) 提高模型的准确性和泛化能力

残差模块可以提高模型的准确性和泛化能力。由于跨层连接可以直接传递梯度信息,使得模型更加容易收敛和优化。此外,残差模块可以避免信息的丢失和变形,从而提高模型的表达能力和泛化能力。

(3) 加速模型的训练和推理

由于残差模块可以解决梯度消失和梯度爆炸问题,并提高模型的准确性和泛化能力。这可以使深度神经网络在更短的时间内达到更好的性能和效果。

残差模块是一种通用的模块,适用于各种深度神经网络。无论是卷积神经网络

络、循环神经网络还是全连接神经网络，都可以通过引入残差模块来解决梯度消失和梯度爆炸问题，并提高模型的准确性和泛化能力

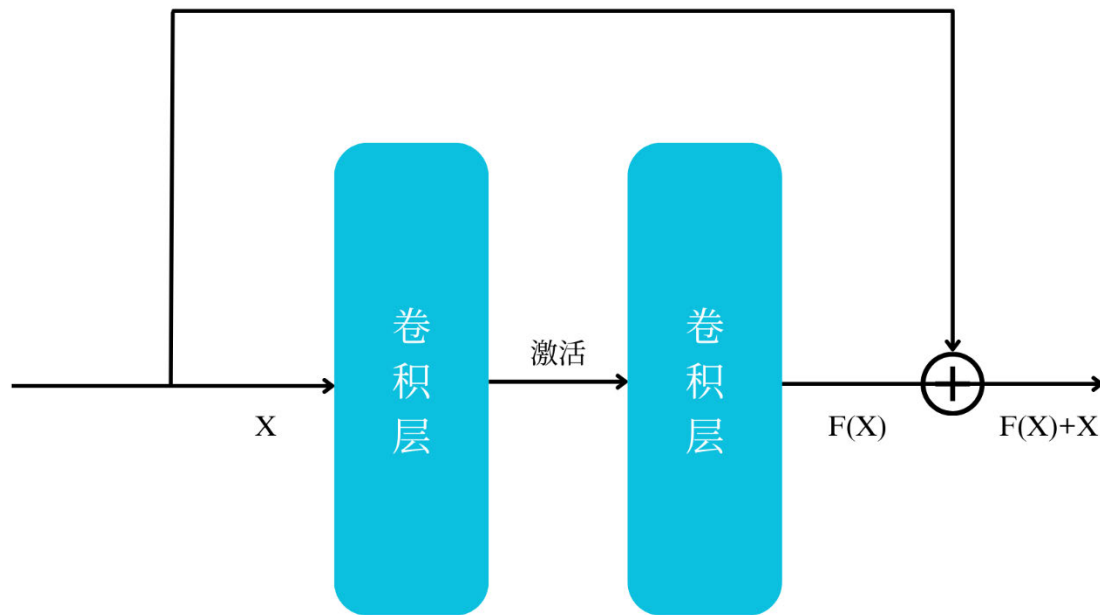


图 3-1 残差结构示意图

3.1.2 P-ResNet 网络结构

本研究中，我们结合心音异常检测的任务特点，搭建了基于残差结构的 P-ResNet 网络框架，网络的结构如图 3-2 所示。我们将卷积神经网络应用到端到端的深度学习任务中，网络将心音信号的时间序列作为输入，并输出预测标签（标签“0”认为心音信号中不存在杂音，标签“1”认为心音信号中存在杂音）。

P-ResNet 模型主体架构由 33 个卷积层以及最后的全连接层和 Softmax 层组成。在 16 个残差模块中，每个模块内存在两个卷积层，卷积层卷积核的长度均为 16，滤波器的通道数为 $32x$ ，其中 x 从 1 开始，每经过四个残差模块 x 变为原来的两倍。在经过第奇数个残差模块（第 1, 3, 5, 7, \dots , 15 残差块）时，输出的长度均保持不变。在经过前四个偶数残差模块（第 2, 4, 6, 8 残差块），输出长度均变为原来的二分之一；经过第五至七个偶数残差模块（第 10, 12, 14 残差块），输出长度均变为原来的五分之一；经过最后一个偶数残差模块，输出长度保持不变。

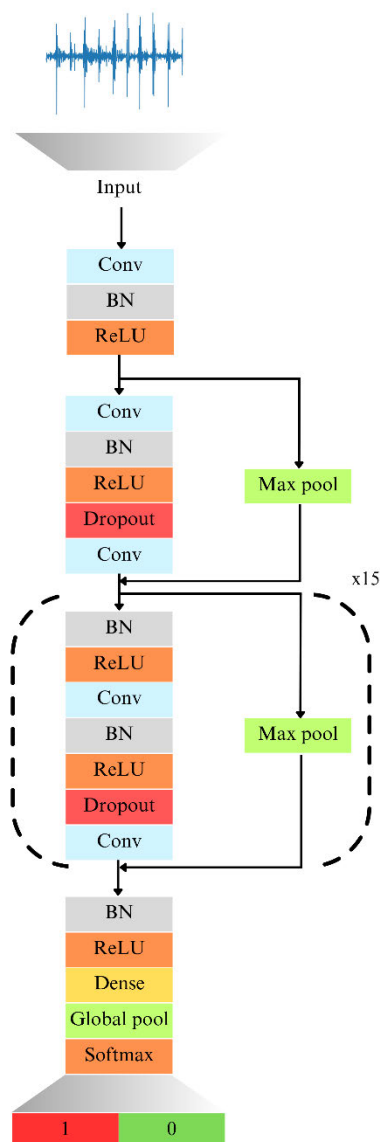


图 3-2 P-ResNet 结构示意图

针对心音信号特点以及 P-ResNet 网络特点，在模型优化过程中，我们选用分类交叉熵（Categorical Cross-entropy）作为 P-ResNet 网络模型的损失函数。分类交叉熵的函数形式如下。

$$L = -\sum(y_i * \log(p_i)) \quad (3-1)$$

其中 y_i 表示真实标签的概率分布， p_i 表示模型预测的概率分布。对于二分类任务， y_i 的值为 0 或 1，表示当前样本属于哪一类； p_i 的值为模型对每一类的概率预测值，通常使用 Softmax 函数将输出值转化为概率分布。

分类交叉熵是深度学习中常用的损失函数之一，主要用于二分类或多分类任

务。它的计算方式基于信息熵的概念，用于度量模型预测的概率分布与真实标签的概率分布之间的差异。具体而言，分类交叉熵将模型预测的概率分布与真实标签的概率分布作为输入，通过对两个概率分布的交叉熵计算得到损失值。交叉熵的值越小，表示模型预测的概率分布与真实标签的概率分布越接近，模型的准确性越高。

3.2 训练和测试

3.2.1 数据划分

如图 3-3 所示，本文在患者水平上采用五折交叉验证方法对数据进行划分。五折交叉验证是一种常见的评估模型性能的方法。在这种方法中，数据集被划分为五个相等的部分，每一部分都被称为一个"fold"，也可称为一折。在五折交叉验证的过程中，我们将进行五次实验。在每次实验中，我们用其中的四折进行训练，剩下的一折用于测试。这样，每折都会成为测试集，确保我们对模型的评估是全面的。五折交叉验证的优势在于它可以在一定程度上减小对数据划分方式的敏感性，使得实验结果更加稳定和可靠，五折交叉验证实现原理见图 3-3。

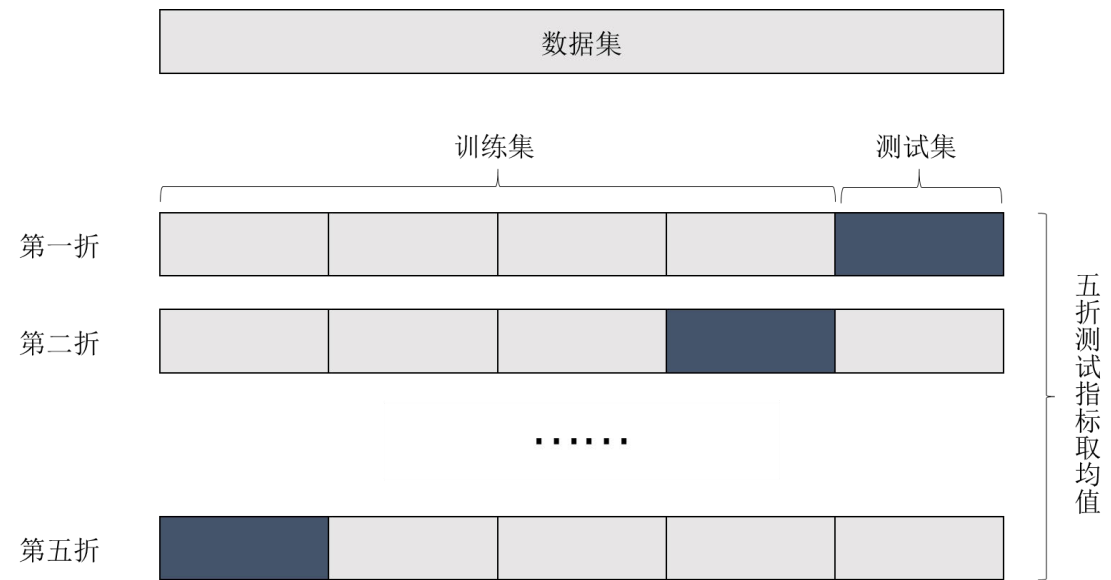


图 3-3 五折交叉验证示意图

3.2.2 模型训练

针对已经经过预处理的心音信号，我们使用五折交叉验证的方法将数据集划分为五折。在每次实验中，我们使用其中四个折作为训练集，剩下的一个作为验证集。

我们采用正态分布初始化器对 P-ResNet 网络进行初始化，并从头开始训练所有的 P-ResNet 网络，在训练过程中，每批次送入网络训练的样本数量设置为 128，选用分类交叉熵作为损失函数用于衡量模型预测与实际标签之间的误差，初始化学习率为 0.001。同时采用早停法监控训练过程，并及时调整学习率。本文采用的早停法的基本原理是在训练过程中监控模型的训练准确度，当训练准确度不再上升时，就停止训练，避免模型过拟合，作为一种简单但有效的正则化方法，防止过拟合，提高模型泛化能力，避免模型在训练集上表现良好但在测试集上表现差的情况。早停法的实现方法有多种，本研究通过以下方法实现早停：

设置一个阈值 n ，当模型的训练准确度连续 n 个 epoch 没有上升时，学习率降为原来的二分之一。

设置一个阈值 m ，当模型的训练准确度连续 m 个 epoch 没有上升时，停止训练。

在每次迭代中，网络会根据训练集中的一个批次 (batch) 的数据进行前向传播，生成预测结果。然后将预测结果与实际标签进行比较，计算损失。接着，我们使用优化器在反向传播中更新网络权重。这个过程会在约 60 个 epoch 中重复进行，直到模型的性能达到一定的稳定水平。

3.2.3 模型测试

如图 3-4，在模型训练完成后中，本文将测试集上的心音信号片段送入模型中进行测试。对于一个完整的心音记录，它在预处理的过程中被分为了多个心音片段，在测试过程中，这些心音片段分别送入保存好的模型之中，对于每一个心音片段，模型对其产生一个预测值。通过计算同一心音记录经预处理产生的多个心音片段预测结果的平均值作为该心音记录的预测结果，从而计算模型在测试集

上的准确率，AUC 等多项指标。

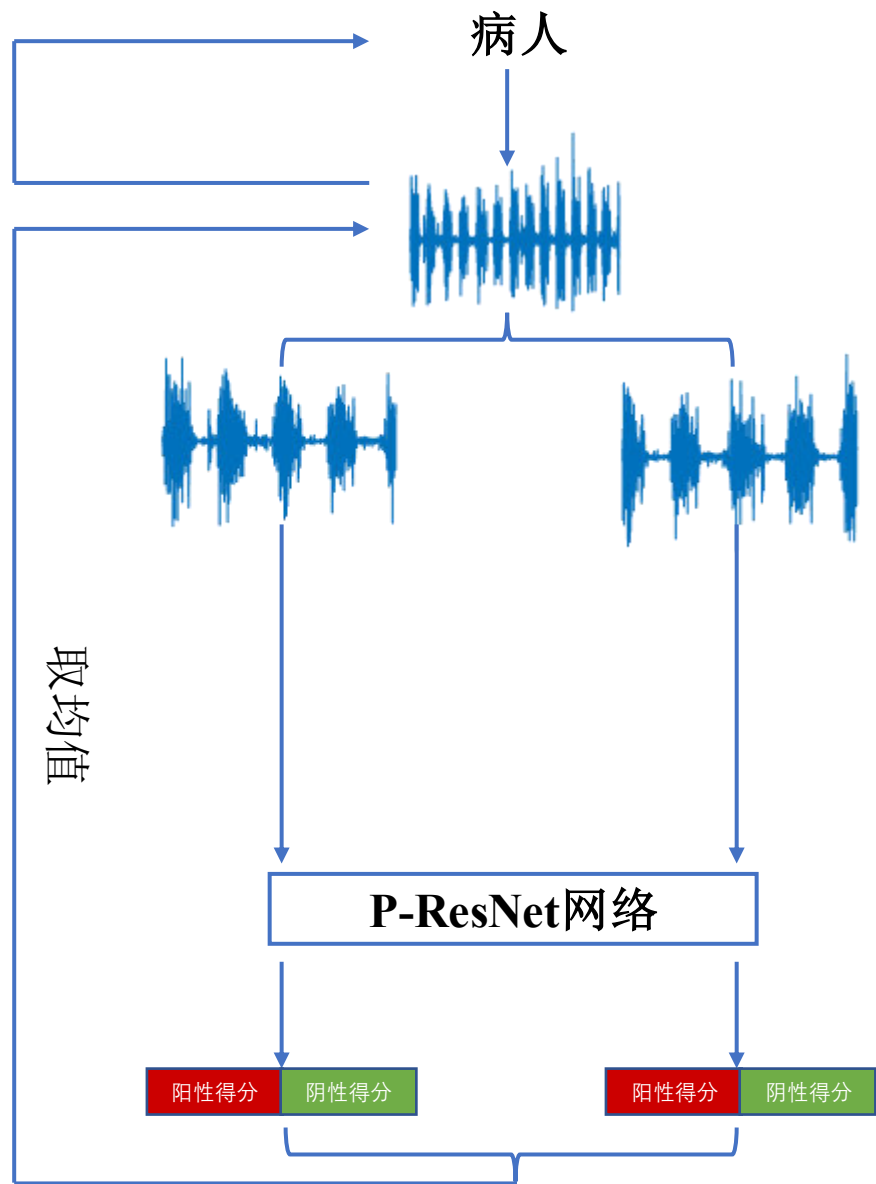


图 3-4 模型测试过程

在五次实验结束后，我们计算模型在五个验证集上的平均性能，作为模型最终的评估结果。通过这种方式，我们可以确保模型在不同数据划分下的表现都得到充分的评估，从而得到更可靠的实验结果。

3.2.4 常见评价指标

在深度学习中，在模型训练完毕之后，将测试集中数据送入模型，根据表 3-1 中的最终分类结果，可以对模型进行评估，常用的指标包括灵敏度，特异率，

平衡准确率，受试者工作特征（ROC）曲线下面积等。

表 3-1 预测结果分类

混淆矩阵		预测值	
		正	负
真实值	正	TP	FN
	负	FP	TN

灵敏度（Sensitivity）：简称为 Sen，灵敏度是指真实正例中被分类器正确识别为正例的比例。在本研究中，灵敏度可以衡量模型对于异常心音信号的检测能力。

$$\text{灵敏度} = \frac{TP}{TP + FN} \quad (3-2)$$

特异率（Specificity）：简称为 Spe，特异率是指真实负例中被分类器正确识别为负例的比例。在本研究中，特异率可以衡量模型对于正常心音信号的检测能力。

$$\text{特异率} = \frac{TN}{TN + FP} \quad (3-3)$$

平衡准确率（Balanced Accuracy）：平衡准确率是指在样本不平衡的情况下，敏感度和特异率的平均值。在本研究中，由于异常样本的数量比正常样本的数量要少得多，因此平衡准确率可以更好地评估模型的性能。

$$\text{平衡准确率} = \frac{\text{灵敏度} + \text{特异率}}{2} \quad (3-4)$$

AUC（曲线下面积）：AUC 是指 ROC 曲线下的面积，通常用于二分类问题。在本研究中，AUC 可以衡量不同概率阈值下模型对于异常信号和正常信号的整体区分能力，是最能准确反映模型分类能力的关键指标。

3.3 实验结果

本研究采用上文中的 P-ResNet 模型，在 George B.Moody PhysioNet 2022 挑战赛公共数据集上进行训练和测试，根据五折交叉验证的原理，将数据集在病人水平随机划分为 abcde 共五折，五折内分别有 10162, 10127, 10209, 9709, 9925 条心音信号片段。每次训练过程中采用其中四折进行训练，一折进行测试，五折交叉验证结果如表：

表 3-2 P-ResNet 网络性能

训练数据	测试数据	灵敏度	特异率	平衡准确率	AUC
abcd	e	73.95%	97.23%	85.59%	0.9115
bcde	a	65.12%	96.11%	80.62%	0.8653
acde	b	76.61%	94.21%	85.41%	0.879
abde	c	75.83%	97.10%	86.47%	0.9585
abce	d	63.25%	95.49%	79.37%	0.8606
		70.95%	96.03%	83.49%	0.8950

表中数据显示，采用 P-ResNet 网络在数据集上 AUC 达到了 0.8950，取得了较好的效果，说明使用 P-ResNet 网络用于心音异常检测和分类是可行的。五折数据之间结果存在一定差异，说明在本研究中采用五折交叉验证更加客观的评估了模型的能力。

3.4 本章小结

本章主要介绍了我们基于残差模块设计的深度学习网络：P-ResNet，同时详细介绍了我们针对心音信号数据设计的测试和训练方法，同时，根据选用的评估指标，展示了模型全方位的性能。通过 P-ResNet 网络框架，我们为后续半监督学习奠定了坚实的基础，将以 P-ResNet 作为核心骨架开展后续半监督学习研究。

4 基于半监督学习的心音异常检测和分类

4.1 半监督方法

平均教师（Mean Teacher）方法是一种半监督学习的方法，它可以有效地利用无标签数据来提高模型的泛化能力和鲁棒性，该方法最初由 Tarvainen 和 Valpola 在 2017 年提出^[39]。与传统的半监督学习方法相比，“Mean Teacher”方法通过使用教师模型来辅助监督，从而减小模型的泛化误差，提高模型的性能。

Zhang 等人通过结合传统的平均教师方法以及更新和锐化方法进一步提出了混合教师方法^[44]。在本研究中，我们采用了该方法，流程图如下所示。

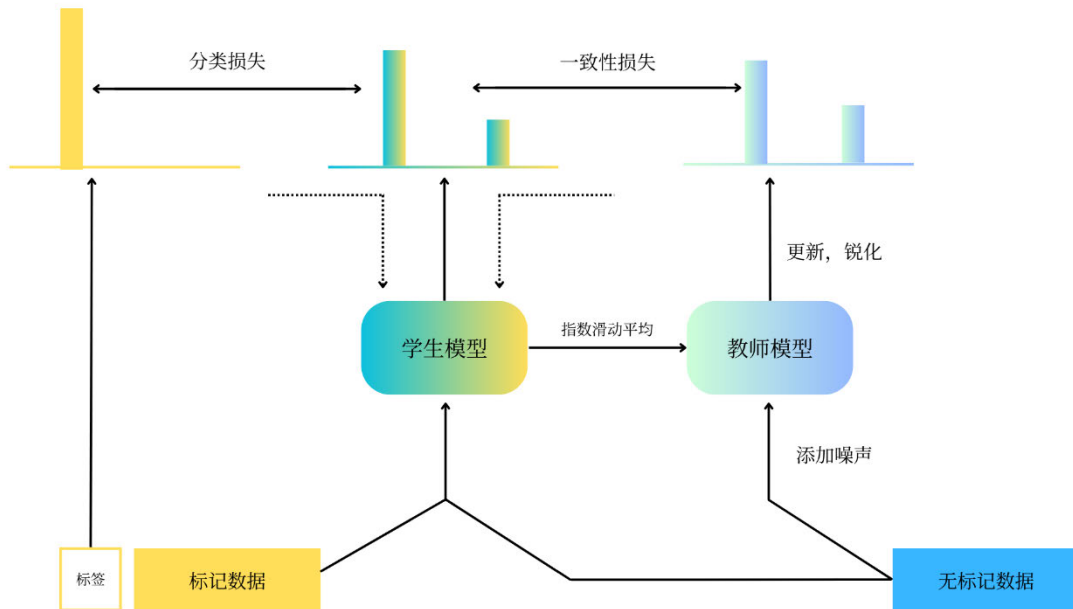


图 4-1 混合教师方法流程图

在我们使用的混合教师方法中，学生模型和教师模型都使用上文中 P-ResNet 网络架构。在每一轮迭代过程中，一定批次大小的带标签数据被送入学生模型，得到对应的预测值 $p_{s,1}$ ；无标签数据根据一定的批次大小被送入学生模型，同样数据经过数据增强后送入教师模型，经过更新锐化处理之后，得到对应的预测值 $p_{s,2}$ 和 $p_{t,2}$ 。然后，计算 $p_{s,1}$ 和真实标签之间的分类损失以及 $p_{s,2}$ 和 $p_{t,2}$ 之间的一致性损失，将两类损失按照一定的权重进行求和之后得到整个模型的总损失，之后

通过该损失来优化学生模型，最后，通过指数滑动平均根据学生模型中参数的值来更新教师模型，至此，一个迭代过程结束，而在一轮训练过程中，往往存在多次类似的迭代过程，伴随着迭代次数的增加，模型性能逐渐趋于稳定，认为此时模型训练完成，最终使用教师模型来用于测试。

在模型优化过程中，分类权重和一致性损失分别反映了监督学习和非监督学习对模型的影响。分类和一致性损失的函数形式如下。

$$\text{分类损失} = w_1 \sum_{i=1}^L y_i \log(p_{s;i}) \quad (4-1)$$

$$\text{一致性损失} = w_2 \sum_{i=1}^L (p_{t;i} - p_{s;i})^2 \quad (4-2)$$

其中 w_1 和 w_2 分别代表分类和一致性损失的权重； L 为类别数，本研究中等等于2； y_i 表示类别 i 的标签； $p_{s;i}$ 表示学生模型预测类别 i 的概率； $p_{t;i}$ 表示教师模型预测类别 i 的概率。一致性损失的权重使用函数 $e^{-w_c(1-x)^2}$, $x = ep/eps$ ，其中 w_c 表示一致性损失权重系数，分类损失权重使用线性函数计算为 $1 - ep - 1/eps$ ，其中 ep 表示当前训练轮数， eps 表示最大总训练轮数。

教师模型的参数在使用 EMA 的每个训练步骤后更新如下，

$$\theta_t = \alpha \theta_t + (1 - \alpha) \theta_s \quad (4-3)$$

其中 θ_t 和 θ_s 分别表示学生模型和教师模型的可训练参数。 α 是一个平滑系数，在前几个 epoch 设置为 0.99，随后设置为 0.999，以帮助于模型可靠地收敛。

在医学信号深度学习中，数据增强方法也是非常重要的技术之一。数据增强可以通过改变信号的形状、幅度、相位等方式来生成新的样本，从而增加数据量并改善数据的质量。其中，加噪方法是一种常用的数据增强方法，它可以模拟真实世界中的噪声，并使模型对噪声具有更好的鲁棒性。

常用的加噪方法包括随机噪声，高斯噪声、泊松噪声、椒盐噪声等。其中，随机噪声是指一种无规律、随机的噪声形式，它可以对信号的特征造成不同程度的影响。随机噪声可以来源于信号的采集、传输、存储等多个环节，例如电子噪

声、热噪声、量化噪声，高斯噪声是加入具有零均值和标准差的高斯分布随机数到信号中，从而模拟信号在传输、采集等过程中所受到的噪声干扰。泊松噪声是在信号中加入泊松分布随机数，模拟信号在光电转换等过程中所受到的噪声干扰。椒盐噪声则是在信号中随机加入椒盐噪声，模拟信号在传输、采集等过程中所受到的不良影响。在本研究中，主要添加随机噪声模仿信号的噪声干扰。

本研究中，我们对教师模型的输出采用更新和锐化方法来进一步泛化性能。本文中采用的更新方法如下所示。

$$Updating(p_{t,w}) = p_{t,w} + \sigma \quad (4-4)$$

其中， $p_{t,w}$ 表示教师模型中弱类别在当前 epoch 的输出概率， σ 表示上一个 epoch 正负样本的平均输出概率之差,在本研究中，设定 $\sigma_{max} = 0.15$ ，即当上一个 epoch 正负样本的平均输出概率之差大于 0.15 时， σ 取 0.15。

在半监督学习方法中，一个常见的假设是低密度分离，这需要分类器的输出具有高置信度而不是模棱两可的。因此，在更新方法之后，我们使用锐化方法来减少“猜测”标签的熵，以加快模型的优化过程^[45-46]，并将更新后的“猜测”标签约束为概率分布，锐化方法被定义为：

$$Sharpening(p_{t,i}) = \frac{p_{t,i}^\gamma}{\sum_{j=1}^L p_{t,j}^\gamma} \quad (4-5)$$

其中 $p_{t,i}$ 代表类别 i 的概率，L 为类别数， γ 为超参数。我们按照^[47]中的建议在模型训练开始时设置 $\gamma=2$ ，并在 20 个时期后设置 $\gamma=5$ 以接近狄拉克（“one-hot”）分布。经过更新和锐化后，每个无标签的样本都会有一个修改后的低熵标签用于训练。

4.2 训练和测试过程

在本研究中，根据五折交叉验证的原理在病人水平将数据集划分为五折，采用混合教师方法对模型进行训练。学生模型和教师模型使用正态分布初始化器进行初始化，监督学习和无监督学习过程中批次样本大小根据数据比例略有不同，，初始化学习率为 0.001，使用早停法监控网络训练结果并及时调整学习率。在本

研究中，主要采用了随机噪声的数据增强方法，以防止模型过拟合并提高泛化能力，在训练过程中，每一轮迭代开始之前，都将带标签数据和无标签数据打乱，防止同一心音记录预处理产生的心音片段集中送入模型从而影响模型学习。训练过程中的每一次迭代都将一定批次大小的带标签样本送入学生模型之中，学生模型对带标签样本产生预测结果之后，与相应的的标签计算分类损失，根据训练轮数计算分类损失权重之后，获得二者的乘积，运用得到的该损失，采用当前学习率对学生模型进行训练。之后将一定批次大小的无标签样本同时送入学生模型和教师模型之中，根据二者产生的预测结果计算一致性损失，同时根据训练轮数计算一致性损失权重，将二者的乘积根据当前学习率训练学生模型，以上步骤完成之后，根据当前迭代批次计算滑动平均参数值，根据学生模型中参数值更新教师模型，至此，训练过程中的一次迭代结束。对于每一轮训练过程，所有无标签数据全部送入模型进行一次训练之后视为一轮训练结束，如果在所有无标签数据送入模型之前带标签数据已经全部送入模型之中，则将带标签数据再次打乱，重新送入模型训练。在每一轮训练结束之后，保存教师模型并选用教师模型对测试集数据进行测试，在经过约 30 轮迭代之后，模型训练结束。测试过程与上文中全监督训练过程类似。

4.3 实验结果

4.3.1 半监督学习方法与全监督学习方法性能对比

我们在病人水平进行五折交叉验证以评估模型性能。在“完全训练”方法中，我们使用了训练集中全部带标签数据；在“部分训练”方法和我们的混合教师方法中，使用训练集中相同的部分带标签数据，此外，我们的混合教师方法还使用了训练集中部分无标签数据。其中“完全训练”方法和“部分训练”方法都属于全监督学习方法。以上三种方法均在相同的测试数据上进行评估。

我们首先评估了使用不同比例的带标签数据的混合教师方法性能，并与“完全训练”方法和“部分训练”方法性能进行了对比。在图 4-2（a）中，混合教师

方法使用了部分带标签数据，剩下的部分数据作为无标签数据使用。图 4-2 (a) 展示了对比结果。可以发现，在使用各种比例的带标签数据的情况下，混合教师方法的性能都相较于使用相同带标签数据的“部分训练”方法有明显提升。由于训练数据减少，“完全训练”方法和“部分训练”方法之间性能下降是十分明显的，而我们的混合教师方法可以极大减小这一下降幅度，表明我们方法可以有效的从无标签数据中学习知识，减小对标注数据的依赖。我们在表 4-1 中详细地展示了混合教师方法在使用 25%带标签数据和 75%无标签数据下的性能。通过比较最关键的指标 AUC 可以发现，我们混合教师方法仅使用四分之一的带标签数据便可实现与使用全部带标签数据的“完全训练”方法相近的性能，证明我们的半监督方法可以极大的减少医生标注工作量，具有更强的临床实用型。

此外，我们进一步探索无标签数据规模对半监督方法性能的影响。我们保持 25%带标签数据不变，从不使用无标签数据开始逐步增加无标签数据数量，直到使用上全部无标签数据，实验结果如图 4-2 (b) 所示。我们可以发现，随着使用的无标签数据增加，混合教师方法模型性能逐渐增加，并逐步逼近使用全部训练数据的“完全训练”方法的性能，表明我们方法具有借助大规模易得的无标签数据实现与“完全训练”方法的可匹敌性能的能力。

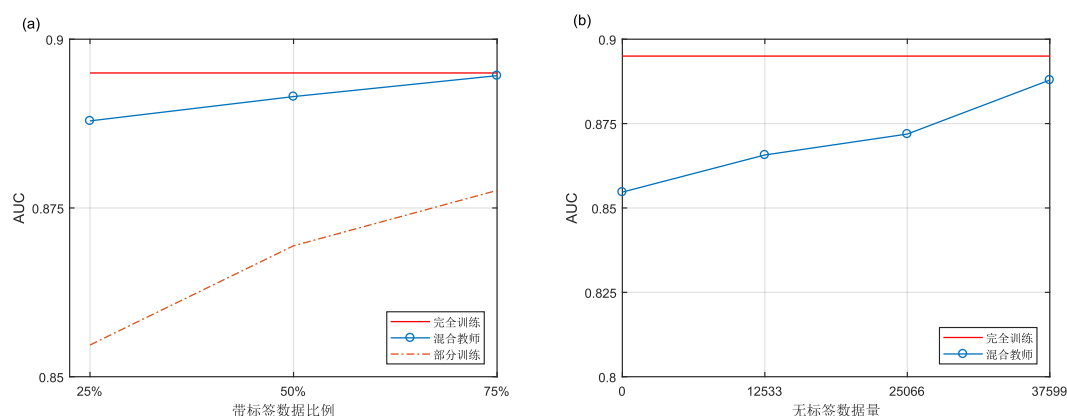


图 4-2 混合教师方法在不同训练数据下的性能

表 4-1 三种方法性能对比

	灵敏度	特异率	平衡准确率	AUC
完全训练	70.95%	96.03%	83.49%	0.8950
混合教师	76.05%	88.36%	82.21%	0.8879
部分训练	65.27%	92.80%	80.03%	0.8547

4.3.2 关键参数与模块分析

在实验过程中，有多项关键参数和模块对于混合教师方法至关重要，直接影响训练方式和模型的各项性能，接下来我们讨论了这些关键的参数和模块。

(1) 一致性损失权重系数 W_C 和最大总训练轮数 eps

为了避免其它模块对消融实验的结果产生影响，因此我们在讨论一致性损失权重和最终截止迭代次数时，移除了数据增强、更新、锐化等模块，实验结果见图 4-3。

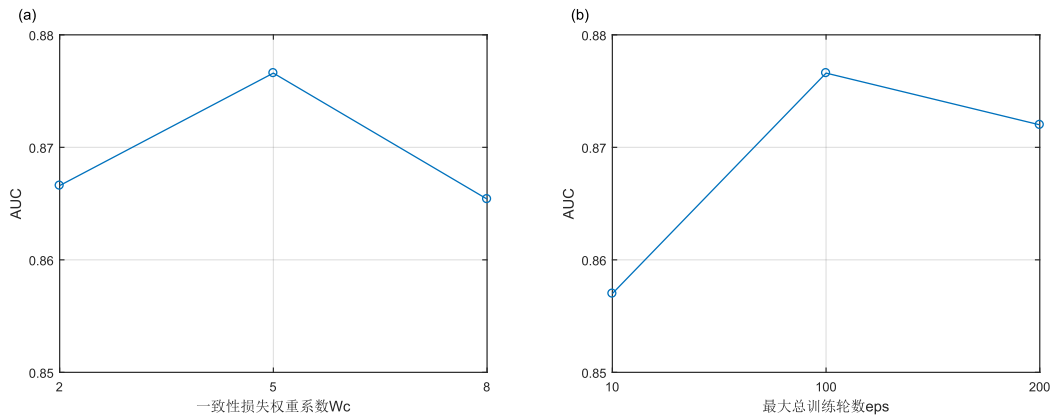


图 4-3 一致性损失权重系数 W_C 和最大总训练轮数 eps 对模型性能的影响

根据实验结果我们可以发现，选用合适的一致性损失权重系数 W_C 以及最大总训练轮数 eps ，对于混合教师模型而言至关重要。这是因为在混合教师模型当中，同时采用无标签和带标签数据对模型进行训练，在训练初期，应该适当的使带标签数据的权重放大，使得模型可以迅速从带标签数据中学习进而快速收敛，如果在训练初期，无标签数据的权重较大，模型将很难收敛。在训练初期结束之后，模型本身具备一定的识别能力，泛化能力，此时再逐渐减小带标签数据的权重，

提高无标签数据的权重，从而使模型整体更多地从无标签数据中学习，从而提高模型的性能，增强模型的鲁棒性和泛化能力。对于一致性损失权重而言，它直接影响了无标签数据在模型训练过程中的重要性，对于截止迭代次数而言，由于分类损失权重，一致性损失权重，都是根据当前迭代次数和最终迭代次数计算而来的，它间接的影响了分类损失权重以及一致性损失权重随着迭代次数的变化规律，从而影响了带标签数据和无标签数据在训练过程中的权重。由此可见，选用适当的一致性损失以及截止迭代次数在混合教师模型训练过程中至关重要。

（2）数据增强

在混合教师方法中，数据增强是至关重要的环节。为了避免其他模块对消融实验产生影响，我们实验过程中去除了更新和锐化模块。数据增强能够模拟现实噪声，同时改变心音信号的幅值，形状等来生成新的样本，改善数据集并能一定程度的增加数据的数量，使得模型能够从新的数据中学习，增强鲁棒性和泛化能力，本文主要通过添加随机噪声来实现数据增强。我们讨论了添加的随机噪声大小对于模型性能的影响，实验结果见图 4-4。

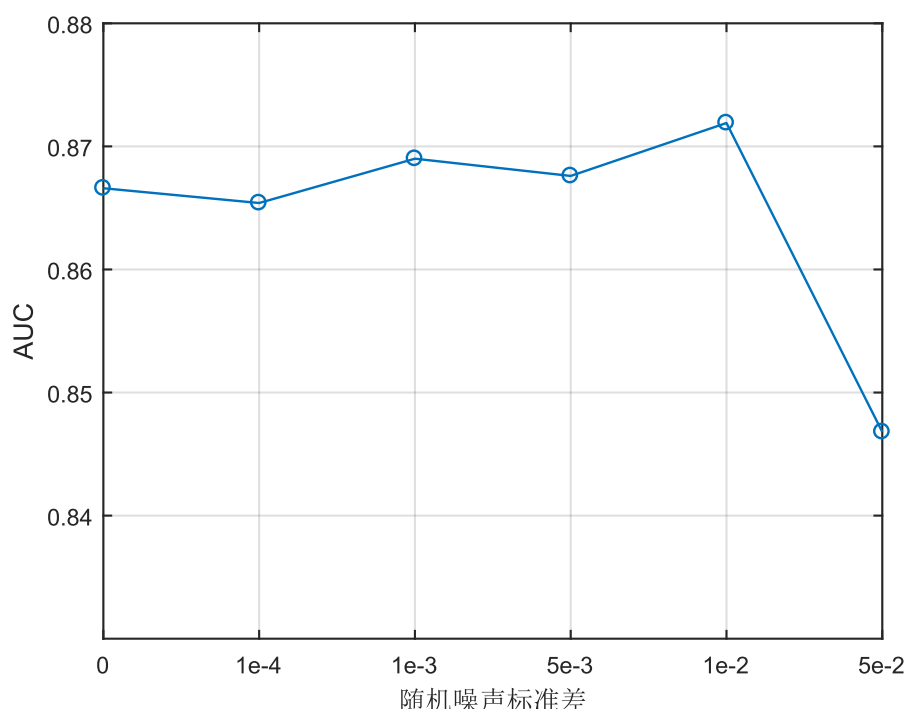


图 4-4 数据增强大小对模型性能的影响

实验结果显示，数据增强大小即添加随机噪声的大小对于模型性能的提升至关重要，当添加的噪声强度较低时，数据增强效果较差，模型无法从加强之后的数据中更多的学习，因此模型效果基本和增强前类似。当添加的噪声强度过大时，原始心音信号被噪声信号淹没，模型无法从失真的信号中提取到有用的信息进行学习，因此，当我们控制噪声强度在一个合适的范围内时，模型可以从加噪的心音信号中学习到更多的信息，增强了模型的泛化能力和鲁棒性。

（3）更新和锐化

在本研究采用的数据集中，我们采用了更新和锐化方法，提升模型的泛化能力，同时使输出具有高置信度，实验结果如表 4-2 所示。

表 4-2 模型性能更新锐化前后对比

	灵敏度	特异率	平衡准确率	AUC
更新锐化前	75.85%	87.02%	81.36%	0.8819
更新锐化后	76.05%	88.36%	82.21%	0.8879

实验结果表明，更新锐化方法在一定程度上使得模型各项性能都有所提升，通过降低预测结果的熵，使得分类器的输出具有高置信度，进而提升模型性能。

4.4 本章小结

本章我们基于 P-ResNet 网络框架，采用了混合教师半监督学习方法，详细地介绍了其使用原理、训练和测试过程，同时我们将本文采用的半监督方法与全监督学习方法对比，我们在仅仅使用四分之一带标签数据的情况下就已经能实现与全监督训练相近的性能。除此之外，我们还针对关键参数和模块进行了分析，证实了一致性损失权重系数，最大总训练轮数，数据增强方法，更新锐化方法对于模型训练的重要意义，进一步证明了混合教师方法的合理性和可行性。

5 总结和展望

5.1 总结

在本文中，我们针对心音异常检测和分类问题展开了研究。本文使用 George B.Moody PhysioNet 2022 挑战赛提供的数据集，首先设计了基于 P-ResNet 的心音异常检测和分类方法，并探究了该全监督学习方法的性能；并进一步以 P-ResNet 作为半监督模型骨干，提出了用于心音异常检测和分类的基于混合教师模型的半监督学习方法，该方法在仅使用部分带标签数据的情况下，取得与使用全部带标签数据的全监督学习方法相近的性能，极大地减少了对标注数据的依赖。

本文的工作总结如下：

（1）对于 George B.Moody PhysioNet 2022 挑战赛提供的数据集，本文首先设计了数据预处理方法，对全部的心音信号进行预处理，为后续深度学习研究奠定数据基础。

（2）经过广泛调研和反复测试，本文在 Resnet 网络基础上设计了适用于心音异常检测和分类的 P-ResNet 网络框架，并通过五折交叉验证全面评估了该全监督学习网络的性能，结果表明模型 AUC 达到 0.8950，实现先进的性能。

（3）为了解决医学深度学习问题中普遍存在的缺少标注数据的问题，本文以 P-ResNet 网络为基础，提出了基于混合教师模型的半监督学习方法。经过一系列的参数探索实验，模型最终在仅使用四分之一带标签数据的情况下，实现了与使用全部带标签数据的全监督学习方法相近的性能，大大减小了模型对带标签数据的需求。

综上，本文为心音信号设计了合适的预处理方法，开发了适用于心音信号检测和分类的深度学习网络框架 P-ResNet，实现了良好的性能。同时设计了用于心音信号分类的半监督学习方法——混合教师模型，极大地减轻了医生数据标注压力，有效地提高了深度学习方法用于心音信号异常检测和分类的临床实用性。

5.2 展望

(1) 在研究过程中,我们发现不同程度的噪声一般对模型训练具有一定的促进效果,可以帮助模型获得更好的性能。在后续研究中,我们可以将经过不同增强方式的数据送入模型之中,设计新的训练方式,进一步提升模型的泛化能力和鲁棒性。

(2) 在模型训练过程中,我们发现对于个别无标签数据,模型的预测能力始终受到一定的限制。我们可以为无标签数据的预测结果开发出置信度模型,挑选出置信度始终较低的数据,帮助进行无标签数据清洗,提高数据集数量,从而进一步提高模型性能。

(3) 根据混合教师模型的特殊性,我们可以改进学生模型和教师模型之间的信息交流结构,进一步帮助模型提高泛化能力和鲁棒性。

致谢

大学四年的帷幕轰轰烈烈地落下，在这段青葱岁月里，一路上我遇上了许多生命中的贵人，值此之际，聊表心意。

感谢我的毕业设计指导老师[]，[]。感谢[]，从去年夏天进实验室以来，手把手教我深度学习，面对我满篇错别字的文章，逐字逐句地帮我审阅批改，实在是太负责太贴心。感谢我的班主任[]，时刻以青年的乐观态度和力量帮助我做好规划和选择。感谢[]。感谢[]，是师生更是兄弟，对我的工作给予充分的帮助和鼓励。感谢[]，是良师，学者风范，让学生受益匪浅，是益友，帮助我度过人生中的艰难时刻。

感谢[]，我们一起挥洒汗水，燃烧激情，镌刻岁月，[]，虽然故事的终点不尽完美，但在我们后面有更多的[]为它续写。感谢同窗四年的室友们，希望[]能早日完成学业，学成归来，建设家乡；希望[]顺利毕业，早日环游世界。感谢这四年里所有志同道合的朋友们，我们在清晨里睡大觉，哦不，是听小鸟唱歌，我们将欢笑洒满[]的各个角落，我们在月色下打比赛，写论文，同赶一份ddl，我们互诉衷肠，共渡难关，我们指点江山，激昂文字，我太爱这段燃烧的岁月。感谢[]，陪哥们儿这么多年，一起经历所有的高峰和低谷。如今大家都已各奔美好前程，祝大家“春风得意马蹄疾，一日看尽长安花”。

感谢我的父母。养育之恩，无以为报，感谢他们给予我全部的支持和信任，让我成为最幸福的人，让我能够自由地为自己选择道路，愿父母平安，健康常在。

一个人无法同时拥有青春和对青春的感受。我拼命地向前奔跑，回过头看大学四年，却怅然若失，恍然如梦。感谢所有陪我度过宝贵时光的可爱的人儿。

天若有情天亦老，人间正道是沧桑。

参考文献

- [1] Organization W H . WHO | Cardiovascular diseases (CVDs) factsheet[J]. World Health Organization, 2017.
- [2] 中国心血管健康与疾病报告编写组, 胡盛寿. 中国心血管健康与疾病报告 2020 概要[J]. 中国循环杂志, 2021, 36(6):25.
- [3] Guyton A C, John E. Hall textbook of medical physiology[J]. Jackson, Mississippi: Saunders Elsevier, 2011.
- [4] Ahlström C. Processing of the Phonocardiographic Signal: methods for the intelligent stethoscope[D]. Institutionen för medicinsk teknik, 2006.
- [5] Deperlioglu O. Heart sound classification with signal instant energy and stacked autoencoder network[J]. Biomedical Signal Processing and Control, 2021, 64: 102211.
- [6] Tuncer T, Dogan S, Tan R S, et al. Application of Petersen graph pattern technique for automated detection of heart valve diseases with PCG signals[J]. Information Sciences, 2021, 565: 91-104.
- [7] Hazeri H, Zarjam P, Azemi G. Classification of normal/abnormal PCG recordings using a time–frequency approach[J]. Analog Integrated Circuits and Signal Processing, 2021, 109(2): 459-465.
- [8] Clifford G D, Liu C, Moody B, et al. Classification of normal/abnormal heart sound recordings: The PhysioNet/Computing in Cardiology Challenge 2016[C]//2016 Computing in cardiology conference (CinC). IEEE, 2016: 609-612.
- [9] Reyna M A, Kiarashi Y, Elola A, et al. Heart murmur detection from phonocardiogram recordings: The george b. moody physionet challenge 2022[C]//2022 Computing in Cardiology (CinC). IEEE, 2022, 498: 1-4.
- [10] Whitaker B M, Suresha P B, Liu C, et al. Combining sparse coding and time-domain features for heart sound classification[J]. Physiological measurement, 2017, 38(8): 1701.

-
- [11] Cornely A K, Mirsky G M. Heart Murmur Detection Using Wavelet Time Scattering and Support Vector Machines[C]//2022 Computing in Cardiology (CinC). IEEE, 2022, 498: 1-4.
- [12] Ren Z, Chang Y, Nguyen T T, et al. A Comprehensive Survey on Heart Sound Analysis in the Deep Learning Era[J]. arXiv preprint arXiv:2301.09362, 2023.
- [13] Ren Z, Cummins N, Pandit V, et al. Learning image-based representations for heart sound classification[C]//Proceedings of the 2018 international conference on digital health. 2018: 143-147.
- [14] Chowdhury T H, Poudel K N, Hu Y. Time-frequency analysis, denoising, compression, segmentation, and classification of PCG signals[J]. IEEE Access, 2020, 8: 160882-160890.
- [15] Xiao B, Xu Y, Bi X, et al. Heart sounds classification using a novel 1-D convolutional neural network with extremely low parameter consumption[J]. Neurocomputing, 2020, 392: 153-159.
- [16] Takezaki S, Kishida K. Construction of cnns for abnormal heart sound detection using data augmentation[C]//Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists. 2021: 18-23.
- [17] Cheng X, Huang J, Li Y, et al. Design and application of a laconic heart sound neural network[J]. IEEE Access, 2019, 7: 124417-124425.
- [18] Alqudah A M, Alquran H, Qasmieh I A. Classification of heart sound short records using bispectrum analysis approach images and deep learning[J]. Network Modeling Analysis in Health Informatics and Bioinformatics, 2020, 9: 1-16.
- [19] Wibawa M S, Maysanjaya I M D, Novianti N K D P, et al. Abnormal heart rhythm detection based on spectrogram of heart sound using convolutional neural network[C]//2018 6th International Conference on Cyber and IT Service Management (CITSM). IEEE, 2018: 1-4.

-
- [20] Ranipa K, Zhu W P, Swamy M N S. Multimodal CNN fusion architecture with multi-features for heart sound classification[C]//2021 IEEE International symposium on circuits and systems (ISCAS). IEEE, 2021: 1-5.
- [21] Balamurugan A, Teo S G, Yang J, et al. ResHNet: spectrograms based efficient heart sounds classification using stacked residual networks[C]//2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI). IEEE, 2019: 1-4.
- [22] Fakhry M, Brery A F A. A comparison study on training optimization algorithms in the biLSTM neural network for classification of PCG signals[C]//2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET). IEEE, 2022: 1-6.
- [23] Qian K, Ren Z, Dong F, et al. Deep wavelets for heart sound classification[C]//2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS). IEEE, 2019: 1-2.
- [24] Dong F, Qian K, Ren Z, et al. Machine listening for heart status monitoring: Introducing and benchmarking hss—the heart sounds shenzhen corpus[J]. IEEE journal of biomedical and health informatics, 2019, 24(7): 2082-2092.
- [25] Zhang W, Han J, Deng S. Abnormal heart sound detection using temporal quasi-periodic features and long short-term memory without segmentation[J]. Biomedical Signal Processing and Control, 2019, 53: 101560.
- [26] Hu Q, Hu J, Yu X, et al. Automatic Heart Sound Classification Using One Dimension Deep Neural Network[C]//Security, Privacy, and Anonymity in Computation, Communication, and Storage: SpaCCS 2020 International Workshops, Nanjing, China, December 18-20, 2020, Proceedings 13. Springer International Publishing, 2021: 200-208.
- [27] Noman F, Ting C M, Salleh S H, et al. Short-segment heart sound classification using an ensemble of deep convolutional neural networks[C]//ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

IEEE, 2019: 1318-1322.

- [28] Krishnan P T, Balasubramanian P, Umapathy S. Automated heart sound classification system from unsegmented phonocardiogram (PCG) using deep neural network[J]. *Physical and Engineering Sciences in Medicine*, 2020, 43: 505-515.
- [29] Avanzato R, Beritelli F. Heart sound multiclass analysis based on raw data and convolutional neural network[J]. *IEEE Sensors Letters*, 2020, 4(12): 1-4.
- [30] Liu K, Yuan L, Huang C, et al. Abnormal Heart Sound Detection by Using Temporal Convolutional Network[C]//2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC). IEEE, 2022: 1026-1029.
- [31] Oh S L, Jahmunah V, Ooi C P, et al. Classification of heart sound signals using a novel deep WaveNet model[J]. *Computer Methods and Programs in Biomedicine*, 2020, 196: 105604.
- [32] Shuvo S B, Ali S N, Swapnil S I, et al. CardioXNet: A novel lightweight deep learning framework for cardiovascular disease classification using heart sound recordings[J]. *IEEE Access*, 2021, 9: 36955-36967.
- [33] Li S, Li F, Tang S, et al. Heart sounds classification based on feature fusion using lightweight neural networks[J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 1-9.
- [34] Chapelle O, Scholkopf B, Zien A. Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews][J]. *IEEE Transactions on Neural Networks*, 2009, 20(3): 542-542.
- [35] Zhu X, Goldberg A B. Introduction to semi-supervised learning[J]. *Synthesis lectures on artificial intelligence and machine learning*, 2009, 3(1): 1-130.
- [36] Nigam K, McCallum A K, Thrun S, et al. Text classification from labeled and unlabeled documents using EM[J]. *Machine learning*, 2000, 39: 103-134.
- [37] Belkin M, Niyogi P, Sindhvani V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples[J]. *Journal of machine*

learning research, 2006, 7(11).

[38] Samuli L, Timo A. Temporal ensembling for semi-supervised learning[C]//International Conference on Learning Representations (ICLR). 2017, 4(5): 6.

[39] Tarvainen A, Valpola H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results[J]. Advances in neural information processing systems, 2017, 30.

[40] Yarowsky, David. "Unsupervised word sense disambiguation rivaling supervised methods." *33rd annual meeting of the association for computational linguistics*. 1995.

[41] Verma V, Kawaguchi K, Lamb A, et al. Interpolation consistency training for semi-supervised learning[J]. Neural Networks, 2022, 145: 90-106.

[42] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[43] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[44] Zhang P, Chen Y, Lin F, et al. Semi-Supervised Learning for Automatic Atrial Fibrillation Detection in 24-Hour Holter Monitoring[J]. IEEE Journal of Biomedical and Health Informatics, 2022, 26(8): 3791-3801.

[45] Grandvalet Y, Bengio Y. Semi-supervised learning by entropy minimization[J]. Advances in neural information processing systems, 2004, 17.

[46] Lee D H. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks[C]//Workshop on challenges in representation learning, ICML. 2013, 3(2): 896.

[47] Berthelot D, Carlini N, Goodfellow I, et al. Mixmatch: A holistic approach to semi-supervised learning[J]. Advances in neural information processing systems, 2019, 32.