

# 第九届全国大学生生物医学工程创新设计竞赛

## 预赛作品报告

### 人工智能驱动的交互可解释性阻塞性睡眠 呼吸暂停专家级诊断系统



作品 ID 号: 7194

参赛学生类型: 研究生

参加赛道: 呼吸系统监测、诊疗器械

组别: 自选项目组

2024 年 6 月

## 内容完整性自查表

说明：

1. 命题组参赛作品请根据命题文件规定或要求的指标填写；
2. 自选组参赛作品请根据实际参赛作品完成情况填写。

完整性类别	任务或技术指标名称	完成效果	呈现方式 (报告中的章节、页码，或者测试报告，或实物、视频等)
参赛作品任务或技术指标要求	1.数据库描述	是	2.2.1 多中心多模态数据集 (P5-P6)
	2.算法结构	是	2.2.2 AI 算法设计与评估分析 (P6-P8)
	3.硬件电路	是	2.2.4 便携式监测设备设计 (P10-P17)
	4.上位机展示	是	2.2.5 交互可解释性 OSA 诊断系统开发 (P20-P23)
	5.演示视频	是	附件
	6.算法诊断性能	是	3.2.1 算法指标及测试结果 (P24-P28)
	7.系统性能验证	是	3.2.2 系统指标及测试结果 (P31-P35)
	8.临床评估	是	3.2.2 系统指标以及测试结果 (P36)
其他			

## 摘要

阻塞性睡眠呼吸暂停（Obstructive sleep apnea, OSA）是一种广泛存在的呼吸疾病，未及时治疗的 OSA 将与众多严重并发症相关。然而，当前的睡眠监测成本过于高昂，依赖于稀缺的多导睡眠图（polysomnography, PSG）设备及睡眠医师的长时分析。因此存在不易普及，智能化分析程度较低的问题，而全自动人工智能（Artificial intelligence, AI）算法由于其黑盒性质和性能瓶颈无法实际应用于临床诊断。

针对这一问题，本项目开发了一种交互可解释性的专家级诊断系统，该系统设计了轻量和多层次的 AI 模型框架并提供了可解释性推理逻辑，良好的交互性设计为医疗专业人员提供了人机协作范式，以快速达到临床睡眠医生的分析性能。此外，系统提供了更灵活的睡眠呼吸多模态数据的监测配置组合，有助于提升患者监测舒适性以扩大筛查规模。系统经过了多中心前瞻性验证，实验结果表明对于 OSA 四类严重程度分级，达到 96% 总体准确率和 95.4% 平均 F1 值的处理时间缩短了约 6.9 倍，而达到 100% 总体准确率和 100% 平均 F1 值的处理时间缩短了约 3.7 倍，凸显了该系统在促进医疗 5.0 体系智能化升级方面的关键作用。

**关键词：**阻塞性睡眠呼吸暂停，交互可解释性，专家级诊断系统，多中心验证，医疗 5.0 体系

# 目 录

1 作品概述.....	1
1.1 背景及意义.....	1
1.1.1 作品背景.....	1
1.1.2 需求分析.....	2
1.1.3 研究目标.....	3
2 作品方案设计及实现.....	4
2.1 作品实现的技术路线概述.....	4
2.2 作品方案设计以及实现过程.....	4
2.2.1 多中心多模态数据集.....	5
2.2.2 AI 算法设计与评估分析 .....	6
2.2.3 灵活简化模态配置探索.....	8
2.2.4 便携式监测设备设计.....	9
2.2.5 交互可解释性 OSA 诊断系统开发.....	18
2.2.6 临床验证与改进.....	21
3 作品测试方案及测试结果.....	23
3.1 测试方案.....	23
3.1.1 算法测试方案.....	23
3.1.2 系统测试方案.....	24
3.2 技术、功能指标及测试结果.....	24
3.2.1 算法指标及测试结果.....	24
3.2.2 系统指标及测试结果.....	30
3.3 技术可行性分析及创新说明.....	37
3.3.1 技术可行性分析.....	37
3.3.2 创新说明.....	39
4 总结.....	43
4.1 作品展示.....	43
4.2 展望.....	44

## 1 作品概述

### 1.1 背景及意义

#### 1.1.1 作品背景

阻塞性睡眠呼吸暂停 (Obstructive sleep apnea, OSA) 是一种常见的睡眠障碍，其特征是在睡眠期间反复出现上气道的阻塞，导致呼吸暂停或显著减少。据估计，全球约有 9.03 亿至 9.7 亿 30 至 69 岁的成年人受到 OSA 的影响。OSA 不仅影响患者的生活质量，还与多种严重的健康问题有关，包括日间嗜睡、认知功能下降、情绪障碍、心血管疾病、中风和某些类型的癌症。这些并发症显著增加了患者的医疗风险和医疗成本，而 OSA 的高患病率也导致了显著的社会经济负担。

目前，尽管多导睡眠图 (Polysomnography, PSG) 在实验室环境中进行仍然是诊断疑似 OSA、评估其严重程度和指导治疗决策的金标准，它通过监测多种生理参数来评估患者的睡眠模式和呼吸事件。然而，PSG 需要在睡眠实验室中进行，这不仅增加了患者的不便，还导致了高昂的成本和资源限制，限制了其广泛应用。此外，PSG 的复杂性和对专业技术人员的依赖也限制了其在资源有限地区的可行性。

为了克服 PSG 的限制，家庭睡眠呼吸暂停测试 (Home sleep apnea test, HSAT) 和便携式睡眠监测设备应运而生。这些方法通过简化的监测设备在家庭环境中进行，可以显著降低成本并提高患者的舒适度。HSAT 通常包括对心电图 (Electrocardiogram, ECG)、血氧饱和度 (Peripheral capillary oxygen saturation, SpO<sub>2</sub>) 和呼吸努力等参数的监测，这些参数与 OSA 的诊断密切相关。尽管 HSAT 在诊断准确性上可能略低于 PSG，但其便捷性和成本效益使其成为筛查和初步诊断 OSA 的有力工具。然而，目前使用单模态信号（如 ECG 和 SpO<sub>2</sub> 信号）作为 OSA 诊断替代方法的研究虽然有所成效，但与多模态融合诊断相比，仍缺乏对其定量效益的全面评估。同时，对异常呼吸事件的详细回顾性分析也存在信息瓶颈。

随着人工智能 (Artificial intelligence, AI) 和机器学习技术的发展，深度学习在医疗领域，尤其是 OSA 的诊断中展现出巨大潜力。深度学习算法能够处理和分析大量的生理信号数据，从而提高诊断的准确性和效率。然而，完全自动化的深度学习模型在泛化性能和可解释性方面仍面临挑战。临床医生和患者对“黑箱”模型的不透明性和不可解释性持谨慎态度，这限制了这些模型在临床实践中的应用。因此，如何通过 AI 技术提供可解释的方法并与睡眠医生合作，实现专家级的诊断准确率，显著减少医生的解释时间以释放医疗资源，具有重要的前景。

为了克服这些挑战，本项目设计了一个名为 Apnea Interact Xplainer (AIX) 的交互式可解释系统。AIX 系统的核心优势在于其能够以较低的成本提供模型预测

异常呼吸事件的详尽解释，并允许用户与预测结果进行交互，从而提高诊断的透明度和可接受性，普惠更多用户。

### 1.1.2 需求分析

OSA 的诊断和管理面临着多方面的挑战和需求。为了更好地应对 OSA 对患者健康的威胁，提高诊断效率和准确性，从以下几个方面进行需求分析：

**需求 1：便携式、舒适、低成本的 OSA 监测设备是早期家庭睡眠呼吸暂停筛查、预防的重要手段，有利于 OSA 早期筛查的普及！**

传统的 PSG 测试需要在实验室进行，费用高昂且操作复杂，导致许多患者无法接受全面检查。实验室测试的环境也可能影响患者的自然睡眠状态，进一步降低诊断的准确性。本项目设计的 AIX 系统提供便携式监测方案，允许用户在舒适的家庭环境中使用集成的多模态传感器进行自我监测，无需实验室环境，降低了成本，提升了用户体验并提高了检查的可及性。

**需求 2：AI 技术的可解释性和透明性是临床接受度的重要保障，开发具有高可解释性的 AI 诊断系统是必然趋势！**

现有 AI 诊断系统的“黑箱”性质往往会导致临床医生或者病人难以理解和接受其结果，限制了 AI 技术在临床中的应用。本项目开发可解释的 AI 诊断系统，提供透明的诊断过程和结果，使临床医生能够理解和验证 AI 的预测。通过界面展示异常呼吸事件的原因和细节，提高临床接受度和信任度。

**需求 3：AI 技术存在全自动性能瓶颈，如何确保 AI 诊断的可靠性并利用其作为提升医疗效率的辅助诊断助手至关重要！设计具有良好交互界面的系统极具价值，通过人机交互，提升效率并达到专家级诊断精度！**

本项目设计的 AIX 系统能自动分析来自多种灵活传感器配置的模态信息，设计的低成本便携式多模态采集设备能够同步采集 ECG、SpO<sub>2</sub>、口鼻气流（Flow）、胸腹运动（Chest）、鼾声（Snore）和心率（Pulse rate, PR）等多种生理信号，提供全面的睡眠呼吸状况评估。AIX 系统包含用户友好的交互界面，有助于操作者快速回溯 AI 模型置信度较低的呼吸事件，并对其进行修正以达到专家级诊断精度。系统还可通过细粒度的多模态数据回顾分析，识别个体差异，制定量身定制的治疗方案，提升预后效果。

通过上述需求分析，可以看出在 OSA 的诊断和管理中，睡眠监测设备的便携性、低成本、舒适性、AI 系统可解释性及交互性是关键要素。解决这些问题，能够显著提高 OSA 的早期发现率和诊断准确性，改善患者的健康状况，减少医疗资源的负担。

### 1.1.3 研究目标

本项目开发的 AIX 系统是一种交互式可解释的 AI 工具，通过分析多模态生理信号诊断 OSA。它结合便携设备和深度学习技术，实现专家级诊断准确率，并通过提供详细的解释和交互功能，显著减少医生的解释时间，优化医疗资源，提升诊断效率和患者舒适度。具体研究目标如下：

- (1) AIX 系统的设计强调了可解释性和交互性，通过阐明模型预测的原因，AIX 系统有助于减少 AI 技术的“黑箱”性质，从而增强临床决策的信心。
- (2) 本设计通过搭建硬件采集平台，并将其与所提出的轻量化 OSA 检测算法相结合，经临床验证可以实现 OSA 的准确分级。
- (3) AIX 系统在提供个别呼吸事件的深入分析方面显示出显著优势，与传统方法相比，AIX 系统能够提供更细致的分析，其粒度控制在与临床标准一致的 30 秒间隔内。
- (4) AIX 系统旨在提供灵活的传感器配置方案，考虑到医疗保健专业人员评估睡眠呼吸事件时主要依赖的信息类型，AIX 系统允许对不同模态配置组进行定量评估，以提供诊断益处的证据。
- (5) AIX 系统使用 AI 技术进行定量评估，以解决不同睡眠专家在视觉筛查评估中的差异，并提供可比较的基准。
- (6) 在临床实践中部署 AIX 系统的结果显示，其在不同简化传感器配置下展现出专家级的诊断性能。这为以患者为中心的个性化诊断和治疗提供了重要潜力，并有助于扩大 OSA 诊断的规模。
- (7) AIX 系统极大地减少了 AI 技术的不透明性，增强了临床决策信心并显著缩短数据处理时间。这不仅提高了诊断效率，还有助于更智能高效地利用医疗资源。

## 2 作品方案设计及实现

### 2.1 作品实现的技术路线概述

OSA 是一种常见的呼吸障碍，表现为睡眠时上呼吸道反复部分或完全阻塞，导致间歇性低氧血症和睡眠中断。未及时治疗的 OSA 将与高血压、心脏病、中风和糖尿病等多种严重并发症相关。因此，早期准确的诊断对于 OSA 患者的治疗和管理至关重要。然而，传统的 PSG 测试复杂且昂贵，且对专业医师的依赖性强，导致诊断资源紧缺。本项目提出了一种基于 HSAT 设备和 AI 技术的诊断系统，旨在开发一种新颖的、高度智能化的医疗诊断系统，专门用于 OSA 的诊断。该系统旨在解决现有诊断方法中存在的资源紧缺和性能瓶颈问题。以下是详细的技术路线概述，主要包括以下几个步骤：1) 多中心数据集收集与清洗。2) AI 算法设计与评估分析。3) 灵活简化模态配置探索。4) 便携式监测设备设计。5) 开发交互可解释性 OSA 诊断系统。6) 临床验证与改进。

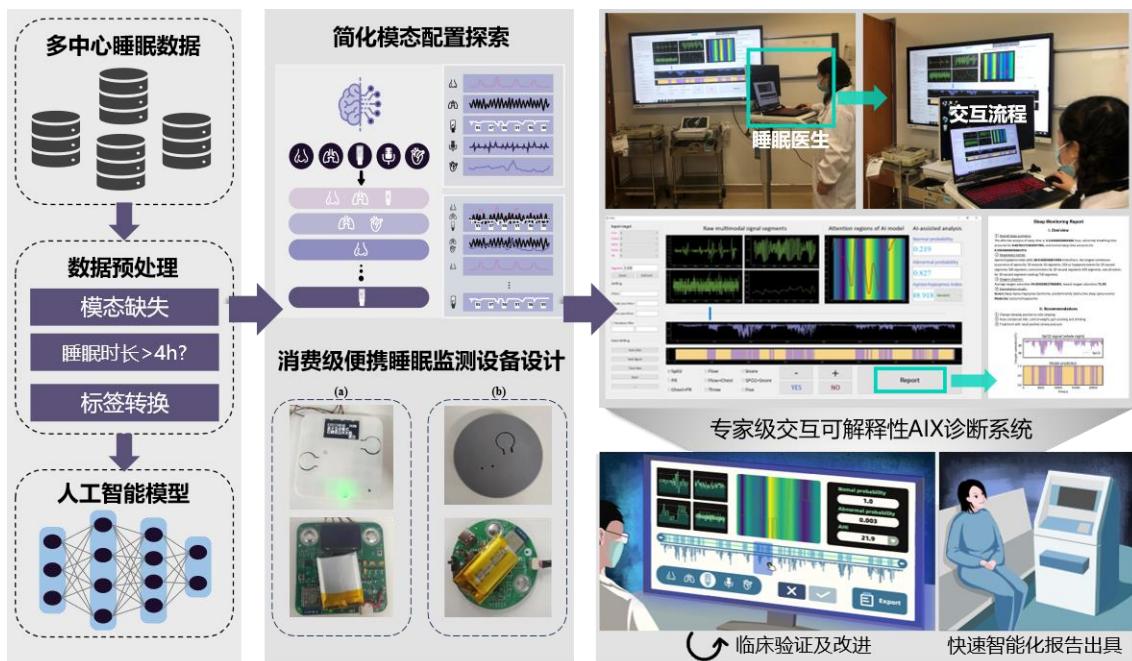


图 2.1 本系统总体技术路线

### 2.2 作品方案设计以及实现过程

本项目搜集了来自三个不同合作医院的多中心多模态睡眠呼吸监测数据集，采集时间从 2020 年 1 月到 2024 年五月。随后本项目通过设计 AI 算法对模态组合进行验证某些简化配置模式在呼吸相关信息不充分条件下对呼吸暂停-低通气指数（Apnea hypopnea index, AHI）评估的准确性。进一步考虑到监测资源的昂贵性，本项目设计了针对两个不同场景下的采集设备，并进一步设计了一个专家级诊断

系统，其具有交互可解释性的界面。最后，本项目在临床环境搜集新的数据，以对系统进行验证和改进。

### 2.2.1 多中心多模态数据集

在开发 OSA 诊断系统的过程中，本项目对数据集的搜集与清洗进行了系统化的设计和实施。首先，多中心数据集的搜集是本项目工作的基础。本项目选择了全国范围内具有代表性的多个医疗中心进行合作，这些中心包括综合医院、专科诊所和社区医疗中心。每个中心的数据集涵盖了不同地域、性别、年龄和健康状况的患者，确保数据的多样性和广泛性。这些数据集包括多种模态的睡眠监测信号，例如脑电图（Electroencephalogram, EEG）、ECG、SpO<sub>2</sub>、Flow、Chest、Snore、PR 等。通过与各中心的紧密合作，我们收集到了大量高质量的原始睡眠数据。

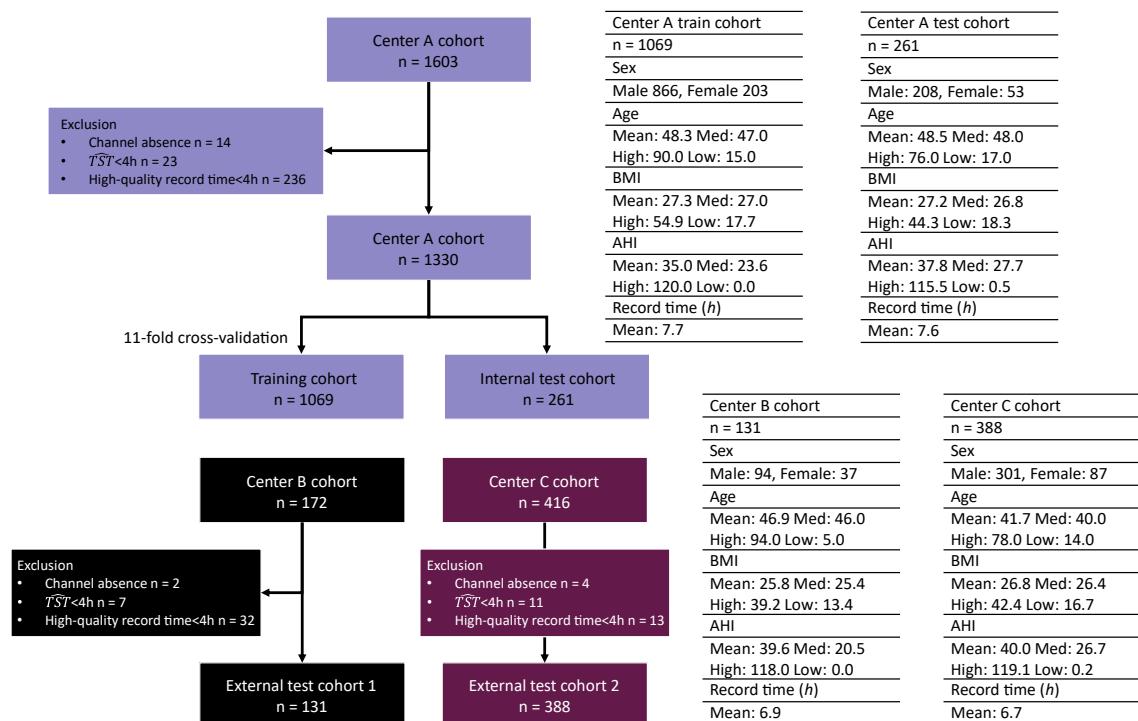


图 2.2 多中心数据集信息

然而，原始数据通常存在许多问题，如噪声干扰、缺失值、血氧伪迹等。因此，数据清洗是确保数据质量的重要步骤。我们首先对数据进行噪声滤波处理，去除数据中的高频噪声和低频干扰。在数据格式方面，进行了标准化处理，使得不同中心的数据在格式和标注上保持一致。所有数据都按照统一的标准进行了信号幅度的归一化、时间戳的格式化和标签的一致化处理。

数据标注是数据清洗过程中至关重要的一步。邀请了经验丰富的睡眠医生对清洗后的数据进行详细标注，标注内容包括呼吸暂停事件的起止时间、事件类型（如

阻塞性、中枢性和混合性)以及其他相关的睡眠事件(如觉醒、腿动)。为了确保标注的准确性和一致性,医生们严格遵循国际 AASM 标准进行标注,标注后的数据被组织成结构化的数据库,方便后续的分析和模型训练。

最终,我们构建了一个高质量的多中心 OSA 数据集。该数据集包含了多模态的睡眠监测数据,覆盖了不同的患者群体和 OSA 严重程度类型,为 AI 算法的设计和评估提供了坚实的基础。通过这一数据集,我们能够确保所设计的 AI 模型在异质性分布的多中心临床数据中具有良好的泛化能力和鲁棒性。详细构成如图 2.2 和表 2.1 所示。

表 2.1 多中心数据集受试者数量分布

数据集	总人数	正常		轻度		中度		重度	
		人数	比例	人数	比例	人数	比例	人数	比例
中心 A-训练	1069	138	12.9%	256	23.9%	220	20.6%	455	42.6%
中心 A-测试	261	36	13.8%	67	25.7%	42	16.1%	116	44.4%
中心 B-测试	131	23	17.6%	30	22.9%	23	17.6%	55	42.0%
中心 C-测试	388	59	15.2%	73	18.8%	80	20.6%	176	45.4%

## 2.2.2 AI 算法设计与评估分析

为了促进生理信号从局部依赖性向全局依赖性的转变,本项目介绍了一种新型的 LEAF-Net,如图 2.4 所示,由四个不同的阶段组成。

第一阶段和第二阶段采用本项目提出的多尺度特征聚合器(Multi scale feature aggregator, MSFA)块的堆叠来捕获局部依赖性。MSFA 块由两个重要组成部分构成:多头卷积(Multi head convolution, MHC)模块,旨在扩大感受野同时捕获多尺度特征;以及特征聚合(Feature aggregation, FA)模块,旨在有效整合来自不同头的信息,同时保持轻量级架构。在第三阶段,MSFA 块和多头自注意力(Multi-head self-attention, MSA)块的结合被用来实现从局部依赖性到全局依赖性的转变。在最后一个阶段,第四阶段,只使用 MSA 块来高效捕获长距离依赖性。

**(1)MSFA 模块:** MSFA 块是第一阶段和第二阶段的主要成分,包括 MHC 模块、FA 模块和高效通道注意力(ECA)网络。**I: MHC:** 如图 2.4 (d) 所示,MHC 模块将输入通道分成 N 个不同的头,并对每个头应用不同深度的可分离卷积。这一操作不仅有助于有效捕获跨尺度特征以捕获局部细节,还有助于减少参数大小和计算成本。**II: FA:** 如图 2.4 (e) 所示,FA 模块涉及由 MHC 模块生成的多种粒度特征的重新组合和分组。具体来说,它涉及从每个头中

选择一个通道来构建一个组，并通过反向瓶颈结构在每个组内执行上下文特征融合操作，从而增强多尺度特征的多样性。最后，通过逐点卷积方法实现所有特征的信息聚合，以促进 MHC 模块内多个头的信息融合。III: ECA\_Net: ECA 可以被视为 SE-Net (Squeeze-and-Excitation Network) 的改进和优化版本（图 2.4 (f)）。这一增强的主要目标是减少模型的参数大小，使其更加轻量级。具体来说，ECA 采用轻量级策略，通过消除 SE-Net 中存在的全连接层，并用  $1 \times 1$  卷积核替代。这一改进的动机在于认识到卷积层在捕获通道间信息方面的卓越性能。在 FA 模块之后引入 ECA 模块增强了模型中特征的重要性和可区分性，从而提高了其整体性能。

**(2) MSFA\_MSA 模块：**图 2.4 (b) 展示了这个模块的组成，它包括一个 MSFA 模块和一个 MSA 模块。通过交错堆叠 MSFA 和 MSA，实现了从局部信息到全局信息的转变，有效降低了计算复杂性。

图 2.4 展示了轻量级进化注意力融合网络 (LEAF-Net) 的整体结构及其各个组成块。（a）MSHEA 的一般结构；（b）多尺度特征聚合器 (MSFA) \_多头自注意力 (MSA) 模块的特定组合；（c）多尺度特征聚合器 (MSFA) 模块的组成；（d）多头卷积 (MHC)；（e）特征聚合 (FA) 模块；（f）高效通道注意力 (ECA) 的特定结构。这种堆叠过程促进了信息从局部到全局的演变，从而优化了计算效率。

**(3) MSA 模块：**在第四阶段，本项目专门使用 MSA 模块来高效捕获长距离依赖性。表 1 提供了不同参数配置的详细模型参数。模型由 Stem 层和第一阶段至第四阶段组成。Stem 层包括步长为 2 的 1D 卷积、输出维度为 64 (64-d) 的卷积层和批量归一化 (BN) 层。第一阶段至第四阶段每阶段均由两个 MSFA 模块、一个 MSFA\_MSA 模块和一个 MSA 模块组成。“dim 64”表示嵌入维度为 64，“sam.ep.r 2”表示 FA 模块的扩展率为 2。

对于一维生理信号 P，其维度定义为 [B, C, L]，其中 B 代表批量大小，C 代表通道数，L 代表信号长度。在本项目中，输入维度为 [256, 1, 1000]，通过 Stem 层后，维度变为 [256, 250, 64]，通道数减少到 64，长度下采样 4 倍至 250，得到的输出特征图尺寸为  $250 \times 64$ 。随后，在第二阶段，特征图分辨率减半，通道数翻倍，得到的输出特征图尺寸为  $125 \times 128$ 。第三和第四阶段的结构与第二阶段相似，它们的输出特征图尺寸分别为  $63 \times 256$  和  $32 \times 512$ 。最终输出通过 AdaptiveAvgPool1d、Flatten、1-d 层，最后使用 Sigmoid 输出概率。

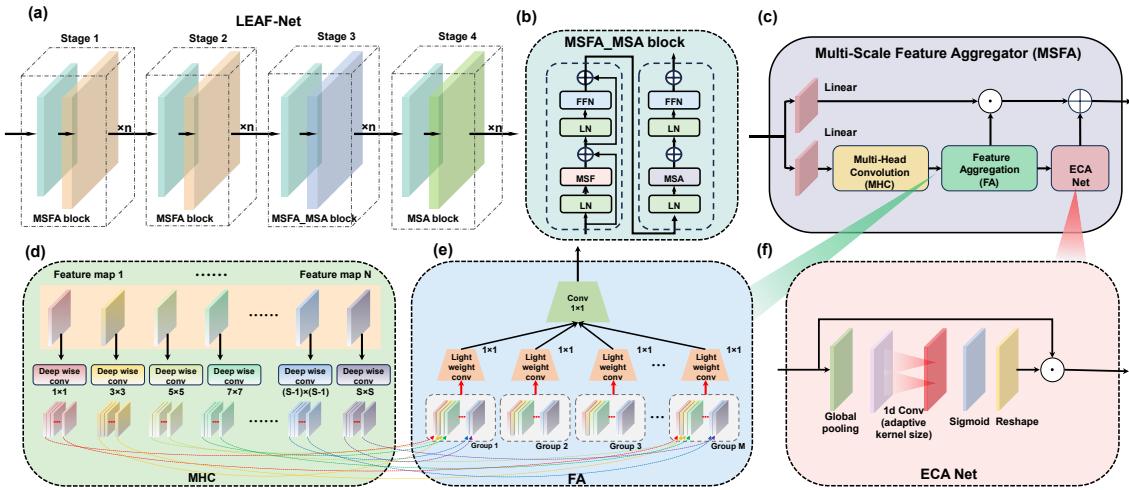


图 2.4 多模态融合进化注意力模型架构

表 2.1 混合模型结构参数

<i>Layer name</i>	<i>Output Size</i>	<i>Detail</i>
<b>Stem</b>	[512, 64, 64]	conv1d 2, 64 – d, BN conv1d 2, 64 – d, LN
<b>Stage 1</b>	[512, 64, 64]	$\begin{bmatrix} \text{dim 64} \\ \text{sam.head 4} \\ \text{sam.ep.r 2} \end{bmatrix} * 2$ conv1d 3, 128-d, LN
<b>Stage 2</b>	[512, 32, 128]	$\begin{bmatrix} \text{dim 128} \\ \text{sam.head. 4} \\ \text{sam.ep. r. 2} \end{bmatrix} * 2$ conv1d 3, 256-d, LN
<b>Stage 3</b>	[512, 16, 256]	$\begin{bmatrix} \text{dim 256} \\ \text{sam.head 4} \\ \text{sam.ep.r 2} \\ \text{msa. head 8} \end{bmatrix} * 8$ conv1d 3, 512-d, LN
<b>Stage 4</b>	[512, 8, 512]	$\begin{bmatrix} \text{dim 512} \\ \text{msa. head 16} \end{bmatrix} * 1$ AdaptiveAvgPool1d, Flatten, 1-d, Sigmoid

### 2.2.3 灵活简化模态配置探索

为了提高 OSA 诊断系统的适用性和用户体验，我们在系统设计中引入了灵活简化模态配置的选择。传统的睡眠监测系统通常需要多个传感器和复杂的设备配置，这不仅增加了患者的负担，还可能影响睡眠质量。因此，该系统致力于简化监

测设备的配置，减少患者的不适感，同时保证诊断结果的准确性和可靠性。首先，本项目对多种睡眠监测模态进行了深入研究，分析各模态信号在 OSA 诊断中的重要性和贡献。通过对大量临床数据的分析，我们确定了几种关键的模态，这些模态在检测呼吸暂停事件和评估睡眠质量方面具有显著的作用。基于这一分析结果，如图 2.5 所示，我们基于生理学解释设计了一套灵活的监测配置方案，重点保留这些关键模态，减少次要模态的使用，从而简化设备配置。

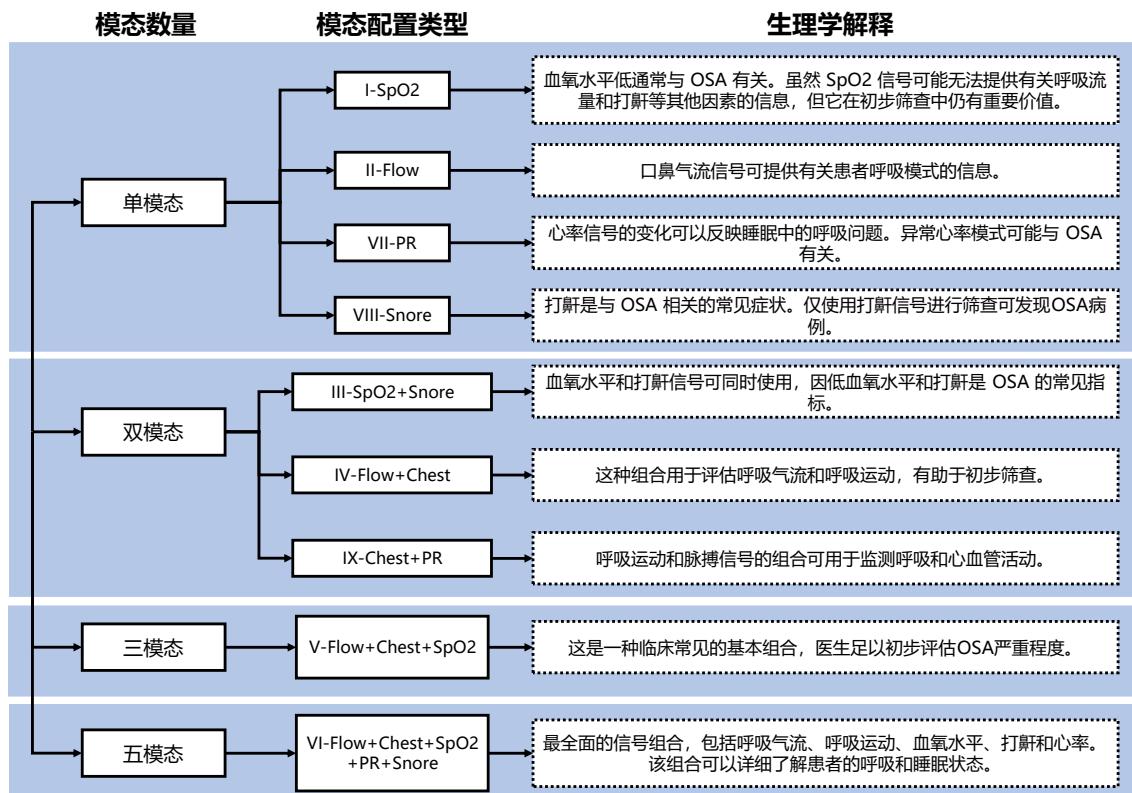


图 2.5 不同模态简化配置生理意义

为了进一步验证简化模态配置的有效性，我们利用多中心数据进行了多种类实验，旨在深入评估不同模态配置对模型性能带来的影响。灵活简化模态配置的探索在提高 OSA 诊断系统的适用性和用户体验方面发挥了重要作用。通过对关键模态的筛选和自适应调整，本项目不仅简化了监测设备的配置，减少了患者的不适感，还保证了诊断结果的准确性和可靠性。这一探索为 OSA 诊断系统的广泛应用提供了有力支持，有助于推动 OSA 诊断技术的普及和发展。

#### 2.2.4 便携式监测设备设计

考虑到 PSG 设备昂贵性限制了 OSA 筛查的范围，为应对 OSA 患者的消费级监测需求，我们提供了不同场景下的便携式监测设备的设计与开发。传统的多导睡眠监测设备通常体积庞大、操作复杂，难以在家庭环境中使用。而便携式设备的设

计旨在提供一种轻便、易用且高效的解决方案，使得患者可以在家中进行长期监测，从而提高诊断的便利性和覆盖面。

在设计过程中，本项目首先考虑了设备的便携性和舒适性。设备的尺寸和重量是关键因素，因此本项目选择了轻质材料和紧凑的设计方案，确保设备在佩戴时不会给患者带来过多的负担。同时，设备采用无线传输技术，减少了传统设备中繁杂的线缆连接，进一步提升了使用的便捷性。本项目还优化了设备的佩戴方式，采用符合人体工程学的设计，使得传感器能够稳固地贴合在患者身体上，同时保证在睡眠过程中不易脱落或引起不适。

为了保证监测数据的准确性和可靠性，本项目在便携式设备中集成了多种高性能传感器。这些传感器包括用于监测血氧饱和度的光电容积脉搏波(PPG)传感器、用于检测鼻/口气流的温度和压力传感器、用于记录胸/腹部呼吸动度的加速度传感器等。各传感器之间通过高精度的同步时钟进行协调，确保数据的同步采集和处理。此外，设备内置了高效的数据处理芯片，能够实时进行数据预处理和初步分析，减轻后端处理的压力。

在数据传输方面，本项目采用了低功耗蓝牙技术，确保数据在传输过程中的稳定性和安全性。为了适应不同的使用场景，设备提供了两种工作模式：实时监测模式和离线记录模式。在实时监测模式下，数据通过蓝牙实时传输到移动端应用或云平台，便于医生和患者随时查看监测结果。在离线记录模式下，数据存储在设备内置的存储器中，患者可以在监测结束后将数据上传到平台进行分析。这种设计不仅提升了设备的灵活性和适用性，还保证了数据的完整性和安全性。

便携式监测设备的设计还注重与 OSA 诊断系统的无缝集成。本项目开发了一款用户友好的移动端应用，患者可以通过手机或平板电脑轻松连接和操作监测设备。应用提供了直观的界面，用户可以查看实时监测数据、设置监测参数并接收诊断报告。应用还具备智能提醒功能，提醒用户按时进行监测并确保数据的及时上传。此外，应用与云平台进行数据同步，医生可以远程访问和分析患者的监测数据，提供专业的诊断意见和治疗建议。通过以上设计，便携式监测设备在提供高质量监测数据的同时，大大提高了患者的使用体验和依从性。患者可以在舒适的家庭环境中进行 OSA 监测，不再受限于传统设备的繁琐操作和不便。这一设计不仅有助于扩大 OSA 筛查的覆盖面，提高早期诊断率，还为 OSA 患者的长期管理和随访提供了可靠的技术支持。

### (1) 系统设计

本项目在深入探讨并提出专家级 OSA 诊断框架的基础上，进一步设计了两种适应不同应用场景的硬件采集方案。方案一（如图 2.6 (b) 所示）是在前序研究工作的基础上进一步发展而来，它采用了一种简化模态的设计思路，专注于采集受试

者的心电脉搏信息。这种设计策略不仅简化了 OSA 监测的复杂性，而且通过集中关注心电脉搏信号，提高了监测的针对性和有效性。方案二（如图 2.6 (a) 所示）则是针对更为专业的医疗场景设计，它旨在通过强化多源生理信息在 OSA 诊断中的作用，从而提供更为全面和精确的诊断依据。在此方案中，除了心电脉搏信号外，还采集了加速度信号以及呼吸信号，这些信号的融合使用，有助于从多角度综合评估受试者的呼吸状况。图 2.7 分别给出了两种方案采集到的信号。

将对方案一的电路系统进行详细阐述。方案一的电路设计以心电脉搏信号的采集为核心，采用了高效的信号放大、滤波以及模数转换技术。电路设计中，首先通过精心设计的电极布局和信号预处理电路，确保心电信号的高质量采集。随后，利用高性能的仪表放大器对心电信号进行初步放大，并配合高通滤波器去除低频噪声和干扰，进而通过精密的模数转换器将模拟信号转换为数字信号，为后续的信号处理和分析打下坚实基础。电路设计中还特别考虑了电源管理、电磁兼容性以及电路的稳定性和可靠性，确保了整个采集系统的高效运行和数据的准确性。此外，为了适应便携式监测的需求，方案一的电路设计还特别强调了小型化和低功耗，使得整个系统便于携带且易于长时间使用。总体而言，方案一的电路系统设计充分考虑了 OSA 监测的具体需求，通过科学合理的电路布局和精心挑选的电子元件，实现了对心电脉搏信号的高效采集和处理，为 OSA 的早期诊断和治疗提供了有力的技术支持下面的小节对方案一的电路系统进行详细阐述。

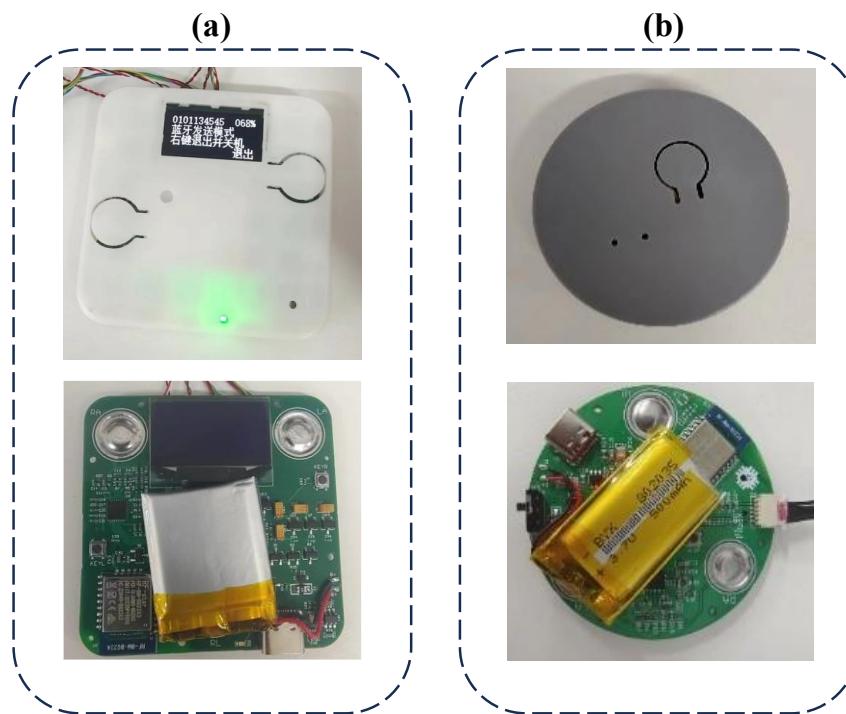


图 2.6 硬件系统实物展示

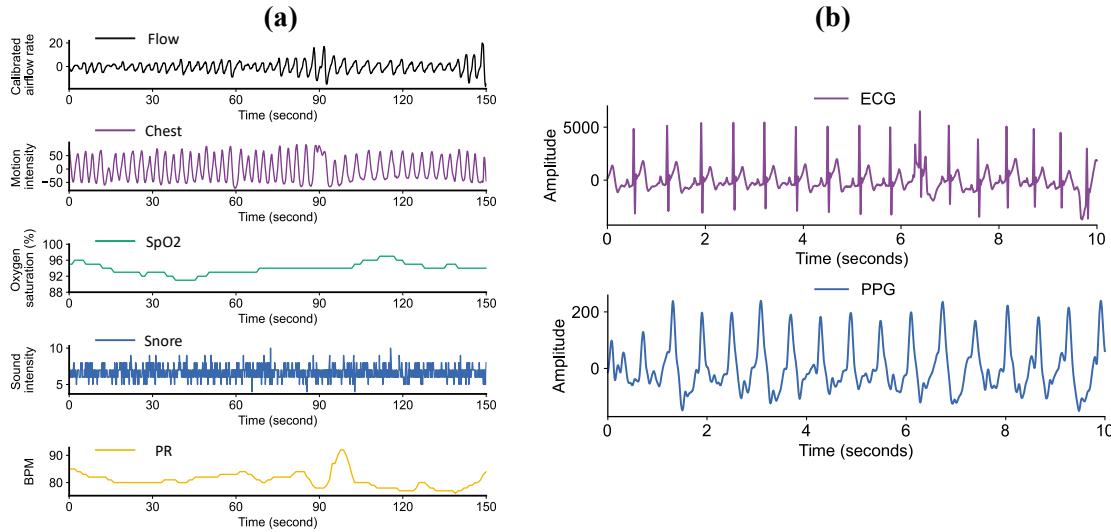


图 2.7 硬件系统采集信号展示

## (2) 主控电路设计

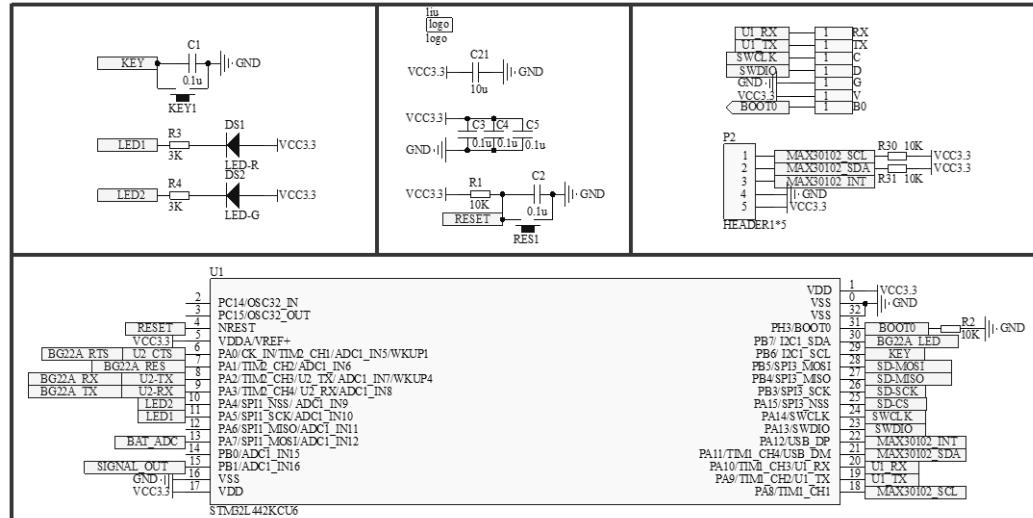


图 2.8 主控电路设计

在本项目中，主控芯片的选择为 STM32L442KCU6，这是一款基于 ARM Cortex-M4 内核的微控制器，具备高性能与低功耗的特点。该器件集成了两个快速 12 位模数转换器（ADC），能够提供高精度的模拟信号转换能力，非常适合用于心电脉搏等生理信号的采集。STM32L442KCU6 的外设接口包括两个按键（Key）和 LED 指示灯。这些外设的使用为用户交互提供了便利，其中，Key1 用于控制蓝牙传输模式的切换，允许用户根据需要选择不同的数据传输方式；Key2 则用于执行主设备的重启操作，便于进行设备的重置和故障排查。LED 指示灯则用于显示设备的电源状态和蓝牙连接状态，为用户提供直观的设备工作信息。具体来说，当设备启动后，LED2 将以快速闪烁的方式持续 60 秒，这一闪烁模式表明设备正处于

蓝牙配对状态。如果在 60 秒内成功配对，LED2 的闪烁模式将变为慢闪，即每 1 秒闪烁一次，表示蓝牙已成功连接并可进行数据传输。若 60 秒内未完成配对，则 LED2 也将转变为慢闪模式，但此时可能需要用户检查配对过程或重启设备。至于芯片的引脚配置，通常会根据具体的应用需求和电路设计进行定制。在本项目中，STM32L442KCU6 的引脚配置将详细定义每个引脚的功能，包括 ADC 输入、按键输入、LED 输出、蓝牙模块接口以及其他必要的通信和控制接口。这些引脚配置将确保整个系统的协调运作，并与外部设备如传感器、显示器和通信模块等实现有效连接。芯片引脚配置如下所示：

表 2.2 引脚配置

引脚	配置	引脚	配置
PA0	蓝牙 RTS 引脚	PA11	MAX30102 SDA 引脚
PA1	蓝牙 RES 引脚	PA12	MAX30102 INT 引脚
PA2	蓝牙 RX 引脚/U2 TX 引脚	PB1	心电电路输出引脚
PA3	蓝牙 TX 引脚/U2 RX 引脚	PB3	TF 卡 SCK 引脚
PA4	LED2	PB4	TF 卡 MISO 引脚
PA5	LED1	PB5	TF 卡 MOSI 引脚
PA8	MAX30102 SCL 引脚	PB7	蓝牙 LED

### (3) 心电采集前端电路设计

心电采集前端的电路设计主要有 0.05Hz 双极点高通滤波器与 45Hz 双极点低通滤波器，滤除心电信号中的基线漂移、运动伪迹、肌电干扰和电极半电池电位等。同时还有一级放大（G=100）和二级放大电路（G=10）。

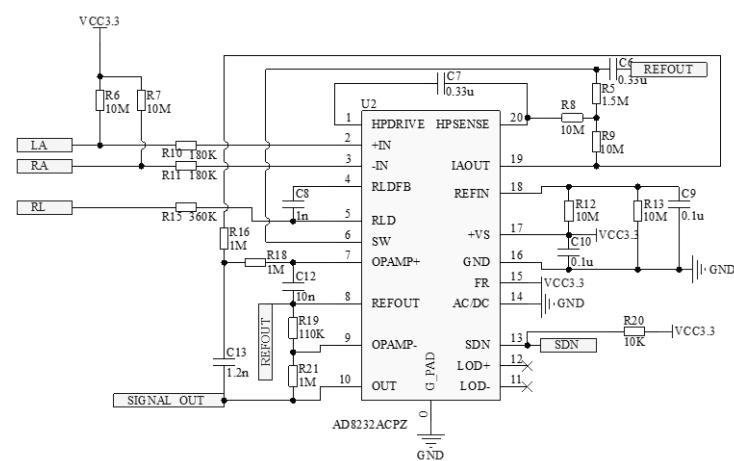


图 2.9 心电采集前端设计

**一级放大和 0.05Hz 双极点高通滤波器设计：**AD8232 是一款高性能的集成电路，其内部集成的仪表放大器（Instrumentation Amplifier, IA）是心电信号放大和滤波的关键组件。该仪表放大器具有固定的增益，通常可将心电信号放大 100 倍，为后续的信号处理提供了足够的信号幅度。心电信号通过电极贴片耦合进入，分别输入到仪表放大器的正向和反向输入端，实现信号的初级放大和高通滤波。在心电信号采集过程中，共模电压变化是影响信号质量的一个重要因素。AD8232 的右腿驱动（Right Leg Drive, RLD）功能可以输出电流至右腿，通过这种方式抵消共模电压变化，从而提高系统的共模抑制比（Common Mode Rejection Ratio, CMRR），确保心电信号的准确性和可靠性。高通滤波器的设计对于去除心电信号中的低频噪声和干扰至关重要。AD8232 配合外围电路可以实现高通滤波，其截止频率的设定通常通过构建 RC 网络来实现。在本设计中，通过实验测量和优化，最终采用了双极点高通滤波器，并在设计中加入了额外的节点，以进一步提升滤波效果。这种设计方式在本电路中相较于其他高通滤波器方案具有更佳的性能，主要表现在滤波器对信号的截止频率计算上。截止频率的计算公式通常为::

$$f_h = \frac{1}{2\pi\sqrt{R_8R_9C_6C_7}} \quad (2.1)$$

由于电阻的误差一般大于电容，所以本设计先确定电阻阻值，如何再计算电容值。在本方案中， $R_8 = R_9 = 10M\Omega$ ，电容值被确定为 $C_6 = C_7 = 0.33\mu F$ 。

**二级放大和低通滤波电路的设计：**本设计中，将低通滤波器设计成较陡的 Sallen-Key 滤波器，该类型滤波器可以更好地从硬件上消除心电信号中夹杂的因肌电干扰和运动伪迹等造成的高频干扰信号，该模块二级放大电路的放大倍数与低通截止频率计算公式如下：

$$G = 1 + \frac{R_{21}}{R_{19}} \quad (2.2)$$

$$f_l = \frac{1}{2\pi\sqrt{R_{16}R_{18}C_{12}C_{13}}} \quad (2.3)$$

将  $G=10$ 、 $R_{22}=1M\Omega$  带入得  $R_{19} \approx 110k\Omega$ 。本项目设计的心电信号采集模拟端系统放大倍数为 1000 倍，心电信号的典型幅值在 1~3mV 之间，放大后的心电信号到达伏特级，满足 A/D 转换要求。在确定外部电路的电容时，首先确定品质因素再计算出电阻电容的大小，为使计算准确方便本项目低通滤波效果的电路设计中取 $R_{16} = R_{18}$ 、 $C_{12} = C_{13}$ ，通过计算得出，其中：选择元件器的阻值为  $1 M\Omega$ ，电容值为  $1.5nF$ ，但是本项目设计的芯片外部增益  $G=10$ ，当增益  $G>3$  时，电路会变的不稳定，解决方案是设置电容  $C_{12}$  容值是  $C_{13}$  的 4 倍以上。为了便于计算，本电路设置 $R_{16} = R_{18} = 1M\Omega$ ， $f=45Hz$  代入公式可得

$$C_{12} * C_{13} = \left( \frac{1}{2\pi * 45 * 10^6} \right)^2 = 12.53 \quad (2.4)$$

因为  $C_{12} > 4C_{13}$  为了方便电容选型, 取  $C_{12} = 10nF, C_{13} = 1.2nF$ 。芯片 REFIN 端的 2 个电阻用于设置中间基准电压, 为了限制分压器的功耗, 本设计的电路设计采用  $10M\Omega$  的大电阻。

#### (4) 脉搏波传感电路

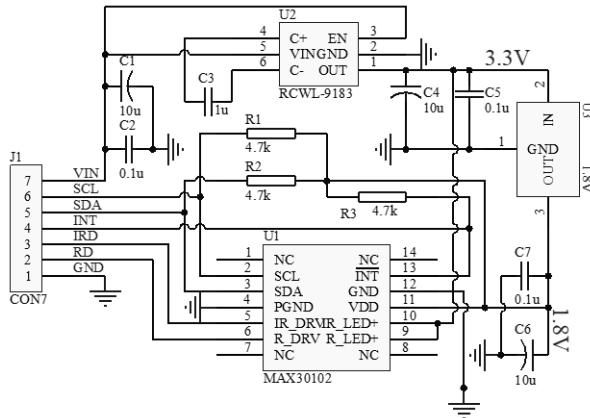


图 210 脉搏波采集电路

本设计中所采用的芯片 MAX30102, 其功能模块的划分体现了高度的集成性和专业性, 主要分为模拟模块、数字模块、通信接口模块以及红外和近红外光源模块。模拟模块的设计包含了环境光消除功能, 以及一个可持续采样的模数转换器 (ADC), 该 ADC 具备 18 位的高分辨率, 能够实现高达 10.24MHz 的采样频率, 并支持 50 至 3200sps 的可编程数据输出率, 这为高精度的信号采集提供了强有力的硬件支持。

数字模块则集成了数字滤波器和数据寄存器, 其中数据寄存器的设计允许通过软件编程进行灵活控制, 从而适应不同的数据处理需求。MAX30102 模块电路的示意图如图 2.10 所示, 其设计简洁而高效。器件的模拟电压输入为 1.8V, 这在整个系统中属于电平供应需求最低的元器件之一。为了与系统的其他部分兼容, 需要通过电平转换技术将电压降至 1.8V。此外, 该模块除了需要 1.8V 的模拟电压输入外, 还需要 3.3V 的电压来对中断引脚和 I2C 通信接口进行上拉处理, 以满足主控芯片的电平需求。MAX30102 的外围电路设计相对简单, 但为了增强系统的稳定性, 根据芯片手册的建议, 在 IR\_LED、R\_LED 与地之间需要加上适当的旁路电容。该芯片共有 14 个引脚, 其中关键的几个引脚包括 pin2、3、5、6、9、10 以及 13。引脚 3 和 2 分别作为通信接口的数据线 (SDA) 和时钟线 (SCL) 引脚, 引脚 5 和 6 用于驱动红光及近红外 LED, 引脚 9 和 10 则是 LED 的供电引脚, 而引脚 13 则用于芯片的中断控制。微处理器通过特定的引脚 PA9 和 PA10 与 MAX30102

芯片的输出端进行连接，实现对双路 PPG 数据的读取。这种设计不仅确保了数据采集的准确性和实时性，同时也体现了系统设计的模块化和可扩展性，为后续的功能拓展和应用开发提供了便利。

### (5) 外设 (BLE/TF 卡/充电电路)

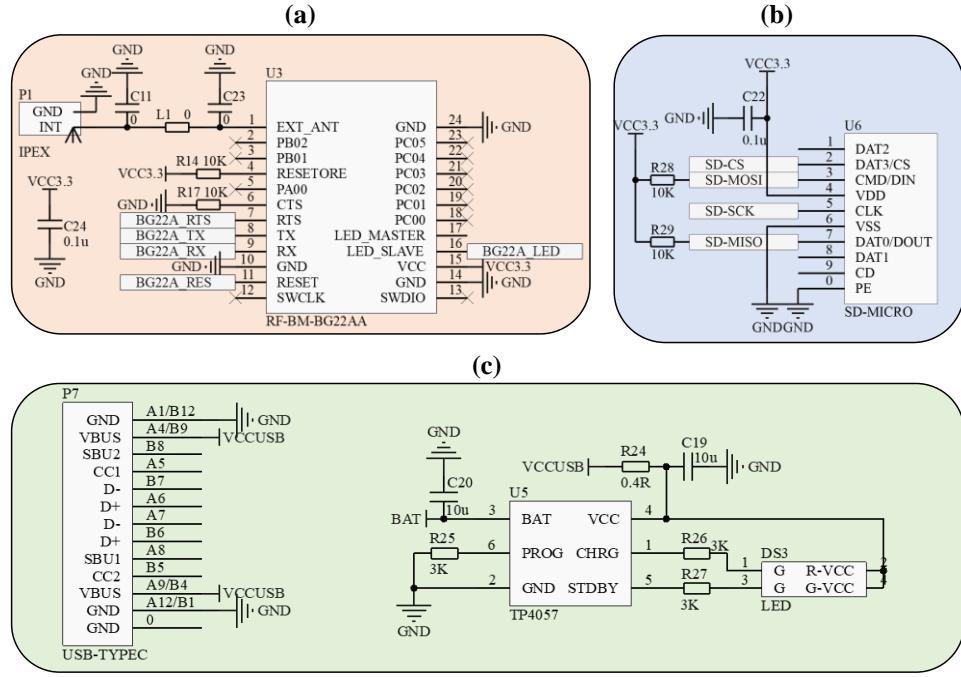


图 2.11 外设电路

本设计在实现基本的生理信号采集功能之外，还集成了多种外设功能，以增强系统的实用性和便捷性。系统电路设计包括了蓝牙低功耗（BLE）模块、TF 卡数据存储模块以及 Type-C 充电模块，这些模块的集成成为系统提供了无线数据传输、数据持久化存储以及便捷的电源管理解决方案。

BLE 模块选用的是支持蓝牙 5.2 协议的 RF-BM-BG22AA 芯片，该芯片具备高性能的无线通信能力，能够实现 ECG 和双路 PPG 信号的稳定传输。芯片的 RTS、TX、RX、RESET 和 LED\_SLAVE 等关键引脚与主控芯片的相应引脚相连，形成了一个高效的通信接口。本设计采用蓝牙转串口的方式，将采集到的心电和 PPG 信号通过 BLE 模块传输至上位机，便于进行实时波形显示和深入分析。

数据存储电路的设计如图 2.11 (b) 所示。TF 卡模块通过 VDD 引脚端连接到稳压模块的 3.3V 输出端，并与 GND 端串联 0.1uF 的贴片电容，以提供稳定的电源和滤除电源噪声。VSS 引脚端接地，CLK 引脚端连接到 MCU 的时钟控制线 PB3 引脚，而 CS、DIN、DOUT 引脚端则分别与 MCU 的 PA15、PB4、PB5 引脚相连，并分别串联 10kΩ 的上拉电阻，确保在无输出状态下信号保持高电平。这样的设计不仅保证了数据存储的可靠性，也便于数据的后续读取和分析。

电源设计电路如图 2.10 (c) 所示，系统采用 Type-C 接口进行充电，提供了一种快速、通用的充电解决方案。此外，系统还配备了 3.3V 稳压芯片 TP4057，这是一款专为单节锂电池设计的恒流/恒压线性充电 IC，具备电池反接保护和防倒充电路，确保了电源的安全性和稳定性。TP4057 的工作电压范围为 4~6V，静态工作电流低至  $40\mu A$ ，其工作电流可通过 6 引脚的电阻进行调节，为系统提供了灵活的电源管理方案。

#### (6) 微处理程序设计

多模态生理信号同步采集系统核心功能在于同步采集心电和光电容积脉搏图等生理信号。该系统通过心电前端模块和 PPG 模块将生理信号转换为模拟数据，随后传输至主控微控制器单元（MCU）进行模数转换（A/D 转换）。转换后的数字信号被 MCU 暂存于内部数组中，为后续的数据存储与处理提供了基础。

系统提供了两种数据存储方案：蓝牙低功耗（BLE）传输和 TF 卡存储。设备启动后，自动进行蓝牙搜索，并通过 LED1 指示灯的快速闪烁来指示搜索状态。一旦与上位机成功匹配，系统将对数据进行打包处理，并通过 BLE 发送至上位机。此外，系统还采用了直接存储器访问（DMA）技术，实现了心电数据与 TF 卡之间的高速传输，显著提升了数据传输和存储的效率，同时减轻了 MCU 的工作负荷，为 CPU 资源的节省和其它功能需求的实现提供了可能。心电贴的采样率由定时器的中断服务程序控制，确保了数据采集的准确性和实时性。A/D 转换后的数字信号通过定时器中断服务程序周期性地以 DMA 方式写入 TF 卡，进一步提高了数据存储的效率和稳定性。

系统启动时，首先进行系统时钟的初始化，确保了系统运行的同步性。随后，系统进行中断优先级的设置，以防止因多个中断同时触发而导致的逻辑错误。在完成延时函数的初始化后，系统依次进行 ADC 模块的初始化，包括采样周期设定、通道模式选择、预分频、扫描模式、数据对齐方式以及触发方式等关键参数的配置。LED 初始化则用于检测 TF 卡是否处于正常工作状态，确保数据存储的可靠性。内存管理初始化负责对内存的分配管理，确保了系统运行的稳定性和效率。SPI 初始化与 DMA 初始化则确定了 SPI 工作模式以及 DMA 数据传输模式和方向，为数据的高速传输提供了保障。定时器初始化则完成了对 TIM2 和 TIM3 的溢出时间、计数模式、时钟分频等参数的设定，为系统的定时控制提供了精确的时序。

在设备作为主设备进入匹配模式时，系统搜索并匹配从设备。本设计提出了两种蓝牙连接方式：蓝牙转串口和蓝牙直连。当选择蓝牙转串口方式时，数据可通过串口以设定的波特率（115200）传输至上位机，实现数据的实时传输和解码。若在 60 秒内未匹配到蓝牙设备，则系统自动切换至 TF 卡模式，进行 TF 卡的初始化。此过程中，系统完成对 TF 卡容量的判断、供电电压的校验以及数据块的写入方式

等一系列操作，确保了数据存储的可靠性和稳定性。最后，在 TF 卡内创建文件对象，并通过定时器中断方式，将 ECG 和 PPG 数据以 DMA 方式写入 TF 卡，实现了数据的高效存储和管理。

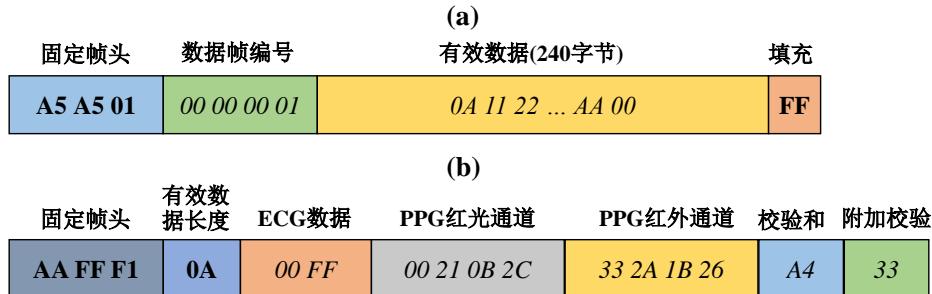


图 2.12 TF 卡数据和蓝牙数据格式

在多模态生理信号同步采集系统中，数据存储格式的规范化设计对于确保数据的完整性、准确性以及后续分析的便捷性至关重要。TF 卡中的数据存储遵循一种特定的结构化格式，具体如下：每个数据帧由帧头、数据帧编号、有效数据以及备用数据四部分组成，总长度固定为 256 字节。帧头由连续的三个字节构成，具有固定的字节序列“A5 A5 01”，用以标识数据帧的起始。紧接着是数据帧编号，采用 uint32 数据类型，为每个数据帧提供唯一标识。有效数据部分占据 240 字节，记录了 30 次独立的采样数据，每次采样由心电（EEG）、RED 和 IR 三个生理信号通道的数据组成。心电数据以 uint16 格式存储，占据前两个字节，RED 通道数据和 IR 通道数据均采用无符号整型格式，分别占据后续的三个字节，且数据存储时高位在前，保证了数据的字节序一致性。在数据帧的最后，有 9 个字节的备用空间，填充为 FF，旨在为数据帧提供结构上的完整性和可能的未来数据扩展预留空间。

对于蓝牙传输的数据格式，其设计同样注重数据的传输效率与准确性。蓝牙数据帧由帧头、有效数据长度、心电通道数据、PPG 的 RED 和 IR 通道数据以及校验和等部分组成，整个数据帧的长度为 16 字节。帧头由四个字节构成，虽然具体内容未明确给出，但其设计目的在于确保数据帧的识别与同步。有效数据长度由单独的一个字节标识，紧随其后的是心电通道的有效数据，以 uint16 格式、低位在前的顺序存储。PPG 的 RED 和 IR 通道数据则分别以 uint32 格式、同样低位在前的顺序存储，各占据四个字节。数据帧的最后包含校验和以及附加校验各一个字节，用以确保数据在无线传输过程中的完整性和正确性，通过校验机制可以有效检测并纠正可能发生的数据错误，增强了数据传输的可靠性。

## 2.2.5 交互可解释性 OSA 诊断系统开发

### (1) 上位机设计

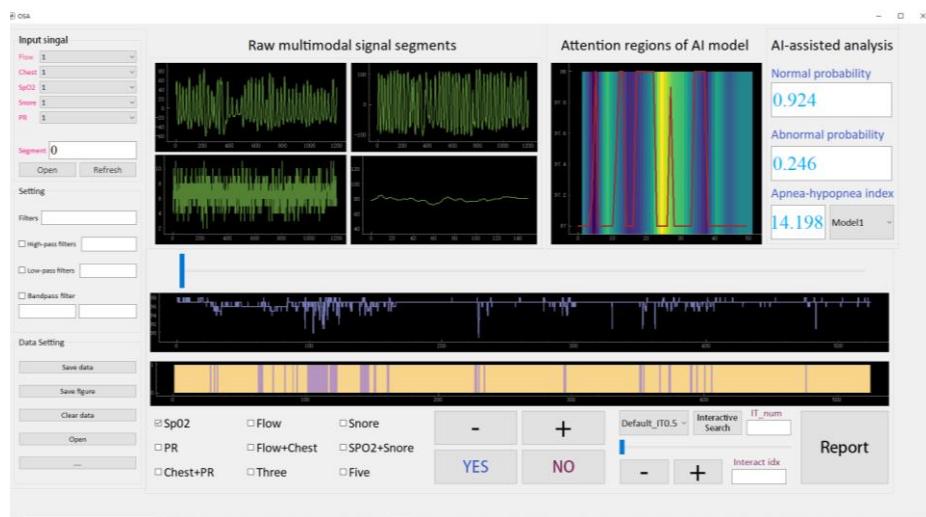


图 2.13 上位机展示

硬件系统将采集的信号传至电脑端，电脑端负责数据的存储。上位机用于显示分析数据，可以对受试者的睡眠状况进行专家级诊断，如图 2.13 是上位机显示界面。使用上位机 PyQt5 进行上位机编程，界面包括串口读取、波形显示、输入模态选择，输入片段粗调和微调，OSA 诊断功能键等。具体功能如下所示：

**模型权重选择功能：**针对四种模态的生理信号，本项目基于 11 折交叉验证的结果，对于每一种模态提供了 11 个权重，当医生认为模型预测结果与所经验结果差距较大时，可以随时切换输入权重；

**输入片段分割功能：**本项目提供两种针对不同场景的输入信号分割方式，级全局调整和局部调整。前者通过滑动条来全局选择所要分析的片段，对于后者可以通过“-”和“+”来逐个分析；

**数据设置：**可以保存四种生理数据；保存波形；以及清除数据等功能；

**波形显示：**本系统提供了 60s 的不同信号的波形展示；

**可解释功能：**本功能可以展示 AI 模型关于血氧饱和度的注意力，可以呈现出与人类专家相同的视觉一致性；

**诊断区：**计算 OSA 的患病概率以及所对应的 AHI 指数；

**交互功能：**本设计提出了一种用于人类专家校准的方案，通过设计的阈值，可以对模型结果不太准确的地方进行微调，并显示具体的位置，从而进一步提高模型准确率。

## (2) 系统可解释性

为了提高进化注意力模型在 OSA 诊断中的可解释性，本项目重点设计了一系列可视化工具和解释机制，使临床医生能够直观地理解和信任模型的预测结果。可解释性是医疗 AI 系统的重要特征，特别是在涉及复杂诊断决策的情况下。

在本项目中，采用了注意力矩阵的可视化方法，通过展示模型在不同时间点对各模态信号的关注程度，帮助医生理解模型的决策过程。注意力矩阵能够显示模型在进行呼吸暂停事件检测时，关注的特征和时间段。例如，当模型检测到一个可能的呼吸暂停事件时，注意力矩阵会高亮显示模型在该时间段内对特定模态信号（如血氧饱和度、鼻/口气流等）的关注度。医生可以通过这些可视化信息，了解模型的决策依据，并验证模型的预测结果。为了实现这一功能，本项目在模型中引入了注意力权重的提取和可视化模块。在每次预测过程中，模型会生成对应的注意力矩阵，记录各个时间点和模态信号的注意力权重。这些权重数据通过可视化界面展示出来，以图形化的形式呈现给临床医生。医生可以通过点击和拖动图形界面，查看不同时间段的注意力矩阵和模型预测结果，深入了解模型的内部工作机制。

此外，本项目还设计了交互式解释系统，提供实时的模型解释和分析功能。交互式解释系统能够实时处理和分析患者的睡眠数据，生成即时的注意力矩阵和预测结果。医生可以通过系统界面查看模型的解释信息，包括各模态信号在不同时间点的注意力分布，以及模型对特定呼吸事件的预测依据。系统还支持医生对模型的预测结果进行反馈，记录医生的修正意见，并在后续的模型训练中进行优化。为了验证和优化模型的可解释性，本项目进行了广泛的临床测试和用户反馈收集。在多个医疗中心的临床测试中，本项目邀请了经验丰富的睡眠医生参与评估，通过对比医生的诊断结果和模型的预测结果，验证模型解释的准确性和实用性。医生们的反馈表明，注意力矩阵的可视化方法能够有效帮助他们理解模型的决策过程，增强了对AI诊断结果的信任和接受度。

### （3）系统可交互性

为了实现OSA诊断系统的高效性和实用性，模型的可交互性是一个关键的设计要素。通过与临床医生的实时交互，系统不仅能够提供即时的诊断结果，还能接受医生的反馈和调整，从而提高诊断的准确性和可靠性。在系统设计中，本项目开发了一套先进的交互式架构，包括数据处理模块、模型预测模块和交互界面模块。数据处理模块负责实时接收和预处理多模态睡眠数据，包括信号滤波、特征提取和数据标准化。预处理后的数据将被输入到模型进行分析和预测。模型预测模块基于进化注意力模型对预处理后的数据进行实时分析，检测呼吸暂停事件，并生成详细的预测结果。预测结果包括呼吸事件的时间段、严重程度和相应的置信度。交互界面模块提供直观的用户界面，展示多模态数据、模型预测结果和解释性信息。交互界面支持医生对预测结果进行查看、调整和反馈，提高诊断的灵活性和准确性。

交互式系统的核心是实现与临床医生的实时交互功能，具体包括实时数据监测、即时预测与解释以及参数调整与反馈。系统能够实时接收和处理患者的睡眠数据，并在界面上动态展示各类信号的变化趋势。医生可以通过界面实时监控患者的呼

吸状况，及时发现异常事件。系统在实时处理数据的同时，提供即时的呼吸事件预测结果。通过结合注意力矩阵，系统能够为每个预测结果提供透明的解释信息，帮助医生理解模型的决策依据。医生可以根据实时监测和预测结果，对系统的参数进行调整，包括信号滤波参数、模型阈值等。系统支持医生对模型的预测结果进行反馈，记录医生的修正意见，并在后续的训练中进行模型优化。为了提高交互式系统的用户体验，本项目特别注重界面的设计和交互功能的优化。界面设计简洁明了，提供丰富的可视化工具，包括时间序列图、热力图和注释标签。通过可视化展示，医生可以直观地查看多模态信号的变化趋势和模型的预测结果，快速做出诊断决策。系统支持多种交互操作方式，包括点击、拖拽、缩放等。医生可以通过点击时间序列图上的特定事件，查看详细的预测信息和解释性数据；通过拖拽调整信号的显示范围；通过缩放功能放大或缩小特定时间段的信号，方便进行细节分析。系统内置便捷的反馈机制，医生可以随时对模型的预测结果进行评价和修正。系统会记录医生的反馈意见，并在后续的训练中进行模型优化，逐步提升模型的准确性和适应性。

为了确保交互式系统在实际临床中的应用效果，本项目进行了广泛的临床测试和评估。多中心临床测试在多个医疗中心进行，邀请不同背景和经验的睡眠医生参与评估。通过对比医生的独立诊断结果和系统的预测结果，验证系统的准确性和实用性。用户反馈收集与分析收集参与测试医生的反馈意见，分析系统在实际应用中的优缺点。重点关注界面的易用性、预测结果的可靠性和解释性信息的准确性，针对反馈意见进行系统优化和改进。持续优化与更新根据临床测试结果和用户反馈，持续优化交互式系统的功能和性能。定期更新系统版本，确保系统在实际应用中的稳定性和高效性，逐步提升诊断效果和用户满意度。

通过上述设计和优化，交互式系统能够为临床医生提供强大的实时交互功能，显著提升 OSA 诊断的效率和准确性。医生可以通过直观的界面查看多模态数据和模型预测结果，灵活调整系统参数并对模型进行反馈，从而实现专家级的诊断性能，为 OSA 患者提供更加精确和可靠的诊断服务。

## 2.2.6 临床验证与改进

临床验证是 OSA 诊断系统开发过程中至关重要的一环，它不仅检验系统在实际应用中的性能和效果，还为系统的优化和改进提供了宝贵的数据和反馈。为了确保系统的可靠性和适用性，本项目设计并实施了一系列严格的临床验证实验，并根据验证结果不断进行系统改进。在临床验证阶段，本项目选择了多家合作医疗中心，覆盖不同地域和医疗水平的医院和诊所。每个中心选取了一批 OSA 患者作为实验对象，患者的选择标准包括年龄、性别、病情严重程度等因素，确保样本的多样性

和代表性。实验过程中，本项目对每位患者进行了长时间的睡眠监测，使用便携式监测设备采集多模态睡眠数据。这些数据不仅用于验证系统的诊断效果，还为系统的优化提供了丰富的实测数据支持。

为了评估 OSA 诊断系统的性能，本项目将系统的诊断结果与睡眠医生的诊断结果进行了详细对比。医生通过传统多导睡眠监测设备和手工标注方法，对每位患者的睡眠数据进行分析和诊断。然后，本项目使用 OSA 诊断系统对相同的数据进行自动分析和诊断，将系统的结果与医生的结果进行对比和分析。对比结果表明，OSA 诊断系统在呼吸暂停事件检测、事件类型识别和整体诊断准确性方面与医生的结果高度一致，显示出卓越的性能和可靠性。

除了对比分析，本项目还邀请医生对系统的使用体验和诊断结果进行评价。医生们普遍认为，系统的操作界面友好，数据展示直观，诊断结果准确可靠。特别是系统的可解释性功能，使得医生能够更好地理解和信任 AI 模型的决策过程，增强了人机协作的效果。医生们的反馈为系统的进一步优化提供了宝贵的意见和建议。基于临床验证的结果和医生的反馈，本项目对 OSA 诊断系统进行了多项改进。首先，在算法层面，本项目对模型的进化注意力机制进行了优化，提高了对复杂睡眠数据的处理能力和对不同患者数据的适应性。其次，在数据处理方面，本项目进一步优化了数据清洗和预处理流程，减少了噪声和干扰对诊断结果的影响。最后，在界面设计方面，本项目根据医生的建议，对系统界面进行了优化调整，增强了交互性和易用性，使得系统更加符合临床使用的需求。

此外，本项目还开展了系统的长期跟踪研究，观察系统在实际应用中的表现和效果。通过对使用系统的患者进行定期随访和监测，本项目积累了大量的实测数据和反馈信息。这些数据不仅验证了系统的长期稳定性和可靠性，还为系统的持续改进提供了重要依据。通过上述临床验证和改进工作，本项目不仅验证了 OSA 诊断系统的高效性和可靠性，还在实践中不断优化和完善了系统。最终，OSA 诊断系统在实际应用中展现出卓越的性能和广泛的适用性，为 OSA 患者提供了高质量的诊断服务，并推动了 OSA 诊断技术的发展和普及。

## 3 作品测试方案及测试结果

### 3.1 测试方案

本测试方案主要可以分为算法的准确性验证以及系统的实时性验证。算法验证需要和当下主流的模型以及现有研究进行参数对比，系统验证需要将算法部署到上位机中，检验模型的实时性，并采集真实数据。

#### 3.1.1 算法测试方案

##### (1) 训练细节

本设计使用 Focal loss 作为损失函数，它旨在减少具有较大样本量的类的权重。在减少了简单样本对损失函数的贡献的同时，而更加关注难以区分样本。其公式如下：

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (3.1)$$

其中  $p_t$  表示预测结果属于真实类别的概率。当  $p_t$  接近 1 时，相应的  $(1-p_t)$  会更小，减少了易分类样本对总损失的影响； $\gamma$  表示可以平稳地调整易分类样本的权重比例。训练的最大迭代次数设置为 100，早停阈值为 15，批次大小为 512。训练期间使用 AdamW 优化器进行优化，学习率从 0.0001 降低到 0.00001（余弦退火算法）。

##### (2) 实验环境

模型的训练环境如下：Python 3.9，Pytorch (Ver.1.13) 与 NVIDIA A100 GPU (80 GB VRAM)。

##### (3) 评估指标

本设计使用七个指标来评价模型的性能：准确性 (Acc, 正确分类的个数与总数的比率)，敏感性 (Sen, 正确分类的阳性样本数与实际阳性样本数的比率)，特异性 (Spe, 正确分类的阴性样本数与实际阴性样本数的比率)，精确度 (Pre, 正确分类的阳性样本数与预测为阳性的样本数的比率)，F1 值 (Sen 和 Ppv 的调和平均值)、AUC (ROC 曲线下面积)，以及  $R^2_{AHI}$  (AHI 预测的 R 平方值)。它们的计算方式如下：

$$Acc(\%) = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (3.2)$$

$$Sen(\%) = \frac{TP}{TP + FN} \times 100\% \quad (3.3)$$

$$Spe(\%) = \frac{TN}{TN + FP} \times 100\% \quad (3.4)$$

$$Pre(\%) = \frac{TP}{TP + FP} \times 100\% \quad (3.5)$$

$$F1 = \frac{2 * Sen * Pre}{Sen + Pre} \quad (3.6)$$

$$AUC = \text{Area} \ (\text{micro ROC curve}) \quad (3.7)$$

$$R^2_{AHI} = 1 - \frac{\sum_{sub} (AHI_{true} - AHI_{pred})^2}{\sum_{sub} (AHI_{true} - AHI_{mean})^2} \quad (3.8)$$

TP、TN、FP 和 FN 分别代表真阳性、真阴性、假阳性和假阴性。

#### (4) 测试内容

算法测试包括二元段级事件分类，患者级 AHI 预测评估，OSA 四分类严重程度评估，不同场景二元筛查性能评估，与先进工作的对比，以及不同风险因素的相关性评估。

### 3.1.2 系统测试方案

为了验证系统的稳定性，本项目从来自三个不同中心的大规模数据集中探索了 9 种不同模式配置的段级呼吸事件检测性能。这有助于揭示各种模态融合配置在呼吸事件检测中带来的可靠收益。经过数据预处理，最终的训练数据集包含 1,397,158 个样本（994,652 个正常呼吸事件，402,506 个呼吸事件），而测试数据集分别来自三个不同中心的数据集。对于每种模态选择，本项目都使用 11 折交叉验证进行模型训练。每折中的划分基于患者级数据收集的顺序，以确保来自同一患者的数据不包括在不同的折中。此外，为了评估模型在不同测试队列中的稳健性，本项目将测试集分为 11 个子集，并评估从 11 个模型中选择的模型，每个模型在这 11 个测试子集上的验证集损失最低。AIX 系统通过集成模型权重形成了用户操作有好的界面，系统测试包含了系统可靠性验证，与专家一致的可解释性视图展示，系统可交互性验证，以及临床系统性能评价。

## 3.2 技术、功能指标及测试结果

### 3.2.1 算法指标及测试结果

#### (1) 段级分类测试

本项目探讨了 9 种不同的模态选择在段级呼吸事件检测中的性能，旨在揭示各种模态融合类型在呼吸事件检测中带来的可靠优势。对于每种模态选择，本项目使用 11 折交叉验证进行模型训练。每个折叠中的划分基于患者级数据收集的顺序，以确保来自同一患者的数据不包括在不同的折叠中。模型预测的典型模式如图 3.1 所示，以 30 秒的粒度输出异常或正常呼吸模式的概率值。

本项目展示了二元呼吸事件与非呼吸事件分类的 ROC 曲线（图 3.2）和 95% 置信区间（Confidence interval, CI），比较了 9 种生理上有意义的衍生模态组合配置。中心 A 内部测试表明类型 I (SpO2) 的 AUROC 为 0.915 (0.893-0.937, 95% CI)，类型 II (Flow) 的 AUROC 为 0.901 (0.879-0.923, 95% CI)，类型 III (SpO2+Snore)

的 AUROC 为 0.927 (0.900-0.954, 95% CI), 类型 IV (Flow+Chest) 的 AUROC 为 0.911 (0.899-0.923, 95% CI), 类型 V (Flow+Chest+SpO<sub>2</sub>) 的 AUROC 为 0.920 (0.899-0.940, 95% CI), 类型 VI (Flow+Chest+SpO<sub>2</sub>+Snore+PR) 的 AUROC 为 0.913 (0.892-0.934, 95% CI), 类型 VII (PR) 的 AUROC 为 0.804 (0.766-0.842, 95% CI), 类型 VIII (Snore) 的 AUROC 为 0.800 (0.750-0.849, 95% CI), 类型 IX (Chest+PR) 的 AUROC 为 0.893 (0.868-0.918, 95% CI)。

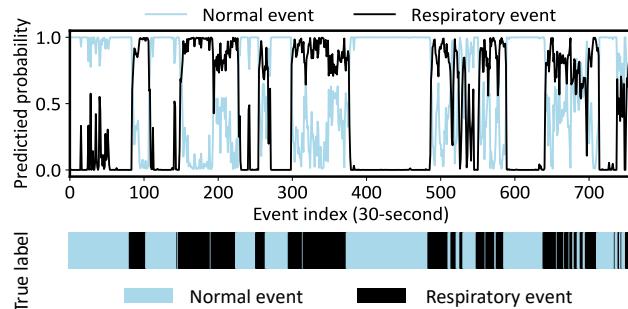


图 3.1 模型预测典型模式

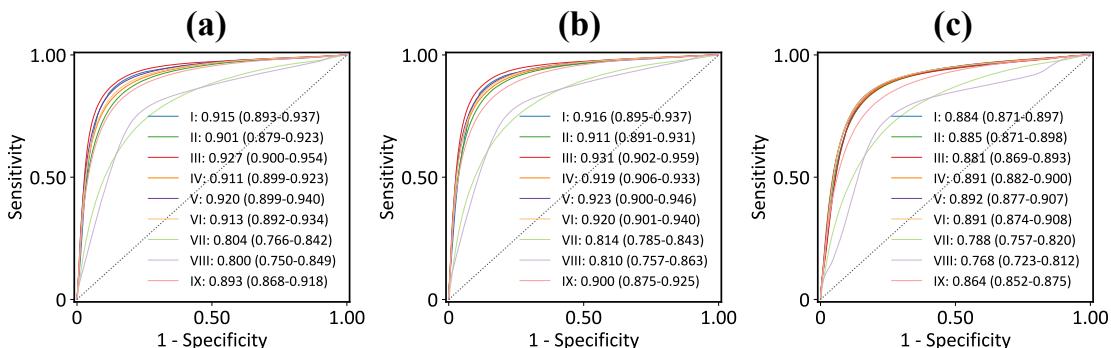


图 3.2 OSA 段级事件分类性能

表 3.1 OSA 段级事件 AUROC

数据集	模态配置类型									
	I	II	III	IV	V	VI	VII	VIII	IX	
中心 A	0.915	0.901	0.927	0.911	0.920	0.913	0.804	0.800	0.893	
中心 B	0.916	0.911	0.931	0.919	0.923	0.920	0.814	0.810	0.900	
中心 C	0.884	0.885	0.881	0.891	0.892	0.891	0.788	0.768	0.864	

研究表明, 类型 I 和类型 II 在模态简化中发挥重要作用, 类型 IX 通过信息互补实现性能提升, 而类型 VII 没有可靠的临床价值。此外, 类型 III 实现了最佳的段级识别性能, 充分体现了 Snore 信号的补充价值。为了进一步探讨结论的可靠性, 本项目在两个外部测试集 (中心 B 队列和中心 C 队列) 上验证了 9 种模态配置, 揭示了类型 III 在不同中心的性能下降更为显著, 表明 Snore 对监测环境的鲁

棒性较低。然而，类型 I、类型 II 和类型 IX 在模态简化方面表现出了很好的价值，与类型 V 和类型 VI 等更全面的模态配置相比，它们的性能下降幅度较小。

## (2) AHI 回归的性能评价

随后，本项目将二元呼吸事件分类结果转换为个体特异性 AHI，并将其与人类专家评分者评估的 AHI 值进行比较。图 3.3 显示了建立多中心 AI 模型的策略以及不同数据集（中心 A 训练集、中心 A 内部测试集、中心 B 外部测试集、中心 C 外部测试集）的患者组成。

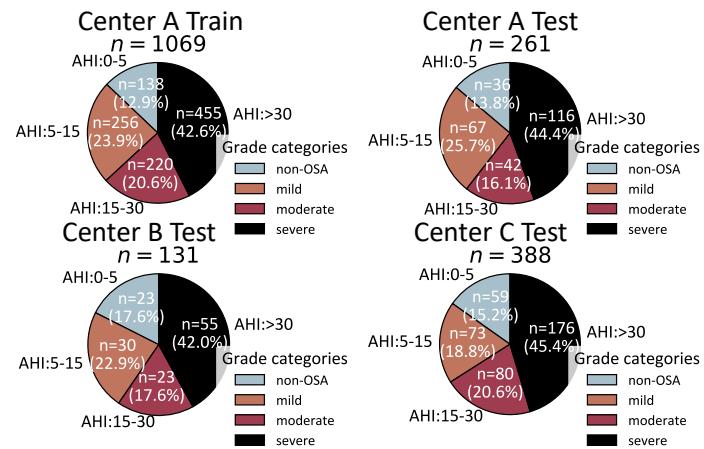


图 3.3 多中心 OSA 患者组成

据观察，OSA 重症的患者往往占很大比例，这进一步强调了早期发现的重要性，因为临床症状通常很明显。图 3.4 显示了多中心不同因素变量分布的小提琴图，其中使用核密度估计计算了来自训练集和测试集的不同受试者的 AHI、年龄 (Age) 和身体肥胖指数 (Body mass index, BMI) 的分布，并截断了负值。

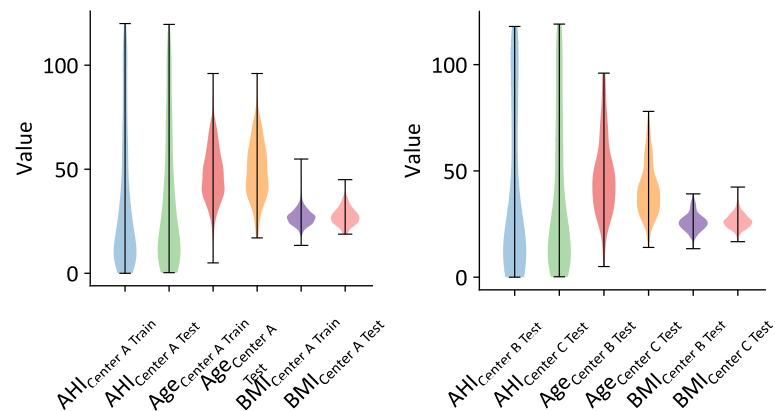


图 3.4 多中心不同因素变量分布

同时本项目展示了中心 A 内部测试集在不同模态配置下进行 AHI 评估的性能

指标，包括 R 平方值、类间相关系数（Interclass correlation coefficients, ICC）和平均绝对误差（Mean absolute errors, MAE），以柱状图的形式呈现，对应于 11 折模型，误差条代表一个标准差（图 3.5）。在评估个体水平 AHI 的实验中，观察到的结果与段级水平事件评估的结果相似。表现前三的是类型 III (SpO<sub>2</sub>+Snore)、类型 I (SpO<sub>2</sub>) 和临床最常用的类型 V (Flow+Chest+SpO<sub>2</sub>)。然而，考虑到模式简化和患者舒适度，类型 II (Flow) 和类型 IX (Chest+PR) 等组合也显示出有希望的性能，为各种患者可接受的方案提供了替代选择。

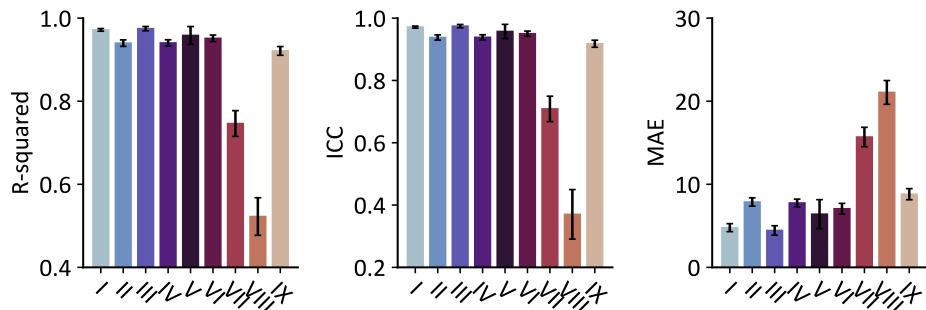


图 3.5 AHI 评估性能

### (3) 不同情况下 OSA 的严重程度分级和二元筛选

图 3.6 和表 3.2 展示了 AIX 全自动预测配置下的分类混淆矩阵和分类性能（类型 I）。此外，本项目考虑了三种分级场景（非 OSA vs. OSA，非轻度 vs. 轻度，非重度 vs. 重度），在中心 A,B,C 三个测试集合中展示了二元分类的总体准确率和平均 F1 值（图 3.7）。本项目全面评估了六种具有优异性能的模态配置（类型 I,II,III,V,VI,和 IX），11 折交叉验证模型的性能范围通过箱线图展示。可以观察的是，类型 V 展示了优异的自动化处理性能，类型 VI 通过引入额外模态信息并未实现显著提升，基于类型 V 简化的单模态类型 I 和 II 同样展现了高性能价值，这一发现强调了 Flow 和 SpO<sub>2</sub> 信息的重要性。类型 V 包含了供医生基于视觉决策的全部信息，这一客观因素蕴含了基于类型 V 进行交互是可靠的，基于类型 I 和 II 的简化无法支持医生进行完备可靠的交互（Flow 中断或减弱但未对应 SpO<sub>2</sub> 下降，SpO<sub>2</sub> 下降但 Flow 未中断或减弱）。然而基于类型 I 和 II 的交互可能更具价值，这不仅促进了监测舒适性的提升，同时对于临床模态缺失情况也能给出有价值的诊断结果，避免由于某个模态缺失或质量不佳带来的重复睡眠监测。

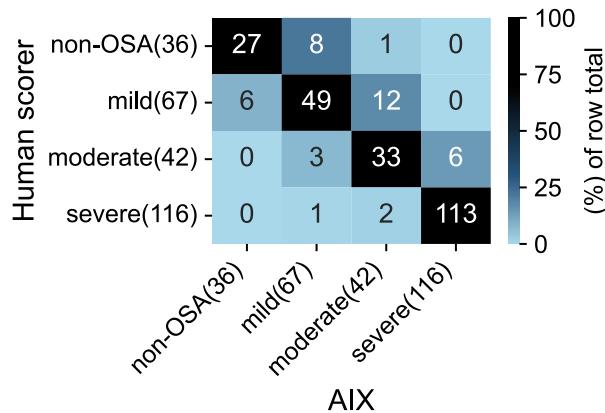


图 3.6 OSA 严重程度分级混淆矩阵

表 3.2 OSA 严重程度分级性能

类别	灵敏度	特异性	阳性预测率	F1 值	总体准确率
正常	0.750	0.973	0.818	0.783	
轻度	0.731	0.938	0.803	0.766	
中度	0.786	0.932	0.688	0.733	
重度	0.974	0.959	0.950	0.962	0.851

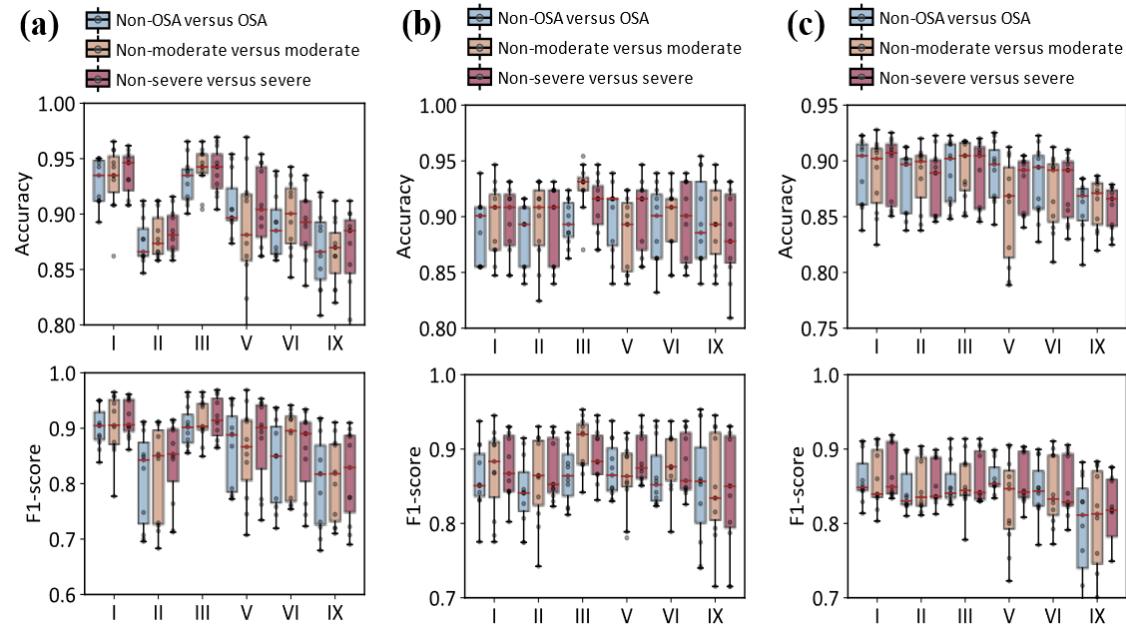


图 3.7 OSA 二元筛查性能

#### (4) 不同深度学习架构和工作的比较

本项目将所提方法与最近最先进的方法进行了比较。值得注意的是，现有的OSA诊断方法主要集中在从单导联ECG中提取呼吸衍生信号或直接使用

SpO<sub>2</sub> 信号。单导联 ECG 原始信号在 HSAT 中往往无法获取，因为原始 ECG 信号通常被以 1 Hz 采样的平均心率所取代，导致缺乏数据的详细细节。这一限制也是使用平均心率诊断 OSA 时表现不佳的原因，因为这些值不能像 RR 间期那样有效地捕捉呼吸障碍的细微变化。对于单模态 SpO<sub>2</sub> 信号，考虑到其提供的有限视觉信息，有必要通过广泛的临床数据来证实其独立的实用性。此外，以往的研究主要依赖于段级事件分类和患者级别的非 OSA versus OSA 预测，在轻/重度 OSA 筛查中表现不全面，而且受试对象数量有限也限制了方法的可推广性。针对这一问题，本项目对大量受试者进行了 OSA 筛查试验，简化了三种诊断方式，提高了患者监护的舒适度。这些包括单模态 SpO<sub>2</sub>，单模态 Flow 和临床常用 Flow+Chest+SpO<sub>2</sub> 组合。它们对二元呼吸事件分类的总体准确率分别为 86.2%、84.8% 和 86.5%，对应的 AUC 值分别为 0.915、0.901 和 0.920，MAE 值分别为 4.99、8.14 和 6.23（基于 11 折交叉验证）。所提出的方法在受试者数量和性能指标方面都优于现有方法（表 3.3）。

表 3.3 不同先进方法对比

Work	Year	Method	Input	Data source	K-fold? (✓ or ✗)	Performance		
						ACC	MAE	AUC
Sharma	2016	SVM	Hermite features of QRS waves (ECG)	PAEDB 70 subs	✗	83.8	-	0.834 (per-segment)
			RRI (ECG)	PAEDB 70 subs	✗	84.7	-	0.869 (per-segment)
Bahrami	2021	LSTM	RRI+RA (ECG)	PAEDB 70 subs	✗	80.2	-	-
			Raw ECG, RA, RRI, RRID (ECG)	PAEDB+UCDDB 95 subs	✗	75.6	9.15	0.800 (per-segment)
Hu	2023	Hybrid Transformer	SpO <sub>2</sub>	OSASUD 30 subs	✗	63.3	-	0.704 (per-second)
			SpO <sub>2</sub>	OSASUD 30 subs	✗	80.0	-	0.889 (per-segment)
Almarshad	2023	Transformer	SpO <sub>2</sub>			86.2	4.99	0.915 (per-segment)
			Flow	Clinical PSG 2191 subs	✓	84.8	8.14	0.901 (per-segment)
Proposed	2024	AIX	Flow+Chest+SpO <sub>2</sub>			86.5	6.23	0.920 (per-segment)

## (5) 不同性别人群 AHI 相关因素分析

考虑到不同性别之间可能存在的 OSA 表型差异，本项目调查了来自中心 A 全

部有效测量数据（男性 1074 名，女性 256 名）的 AHI、BMI 和 Age 变量之间的关联，以了解这些指标在不同性别群体中的关系。这种分析有助于揭示性别对 OSA 患者特征和病理生理特征的影响程度。首先，本项目将数据集分成不同的性别类别，并分别得出男性女性的三个相关性分析组：AHI-BMI、AHI-Age 和 Age-BMI（图 3.8 a, b）。随后，本项目构建了散点图，并计算了各变量对之间的 Spearman 相关系数和  $p$  值。为了更详细地表示变量信息，本项目呈现了所有数据点的概率密度估计。对于 AHI-BMI 和 AHI-Age 分析，不同的颜色（蓝色、浅黄色、红色和黑色）代表非 OSA ( $AHI < 5$ )、轻度 ( $5 \leq AHI < 15$ )、中度 ( $15 \leq AHI < 30$ ) 和重度 ( $AHI \geq 30$ ) 患者。而在 Age-BMI 组分析中，上述颜色分别表示未成年人 ( $Age < 18$ )、年轻成年人 ( $18 \leq Age < 30$ )、中年人 ( $30 \leq Age < 60$ ) 和老年人 ( $Age \geq 60$ )。研究结果表明，在两性中，BMI 和 AHI 显示出相似的相关系数。然而，在 Age 和 AHI 相关性分析中，女性显示出比男性更强的相关性，这表明年龄是女性特别关注的因素。所有相关性都有相应的  $p$  值进行记录，以消除随机性引起的相关性。此外，本项目观察到不同性别之间年龄和 BMI 之间呈现了微弱的负相关性，且该相关性在男性中比女性略大，表明在女性中作为因素相对独立。

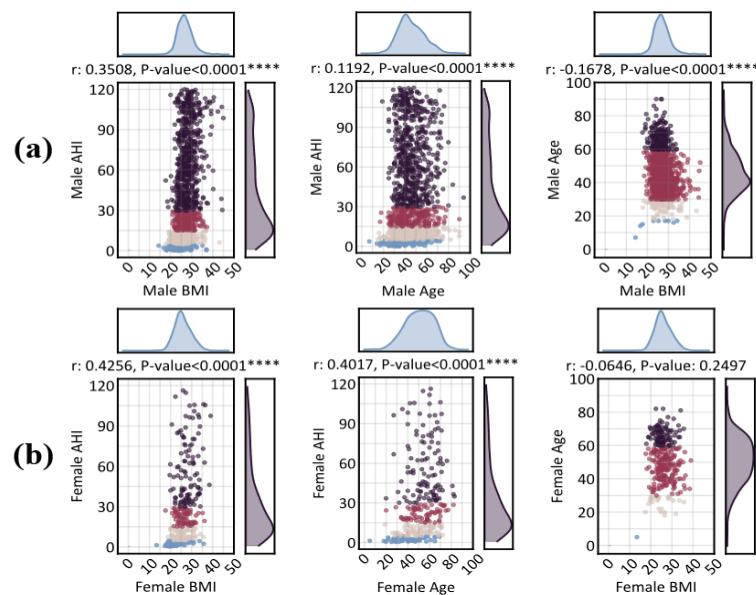


图 3.8 不同性别 AHI 相关因素分析

### 3.2.2 系统指标及测试结果

#### (1) AIX 系统稳定性及可靠性验证

在临床 AI 系统应用的背景下，AI 模型在不同人群中的稳定表现至关重要。因此，本项目研究了具有临床潜力的三种模式类型（类型 I、类型 II 和类型 V）对不同性别（男性和女性）、年龄组（0-40, 40-60, 60 以上）和 BMI 组（0-24.9, 24.9-

29.9, 29.9 以上) 的表现影响 (图 3.9)。本项目综合分析了 AI 模型在中心 A,B,C 所有测试患者中的二元呼吸事件分类性能, 结果表明类型 I 在不同患者组中表现出稳定的检测性能, 而类型 II 和类型 V 在 Age-III 亚组中表现出下降的性能。这归因于老年患者的生理解剖变化和信号采集依从性降低, 使他们更容易在 Flow 信号采集中出现松动或拔掉鼻气管, 从而导致性能表现不稳定。

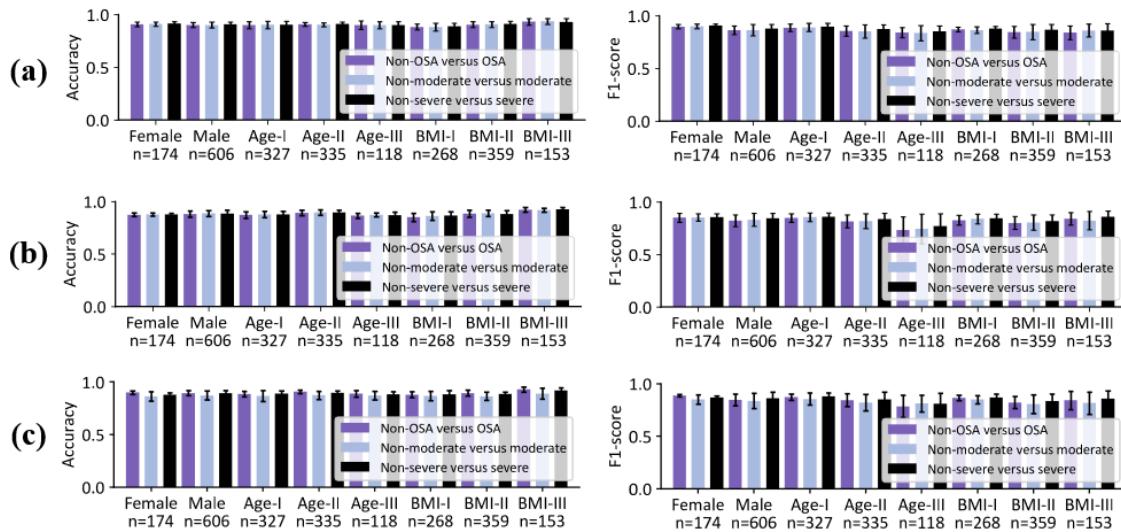


图 3.9 模型稳定性分析 (a-类型 I, b-类型 II, c-类型 V)

此外, 为了展示 AIX 系统在区分不同严重程度 OSA 的稳定性优势, 本项目基于类型 I 对中心 A,B,C 三个测试集合中 (图 3.10) 的预测 AHI 在四个 OSA 类别 (非 OSA, 轻度, 中度, 重度) 上进行了显著性分析 (Two-tailed Mann-Whitney U-tests), 结果表明了 AIX 在内部测试集和多中心外部测试集上的 AHI 预测模式具有一致性, 体现了其泛化性能力。

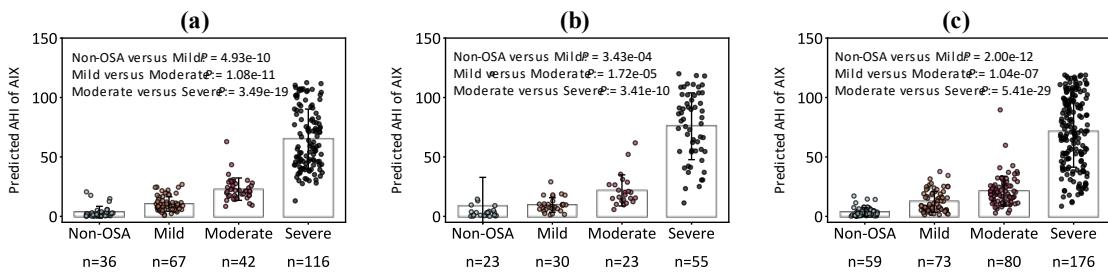


图 3.10 不同中心 AIX 对于严重程度的区分能力 (类型 I)

同时, 为了测试 AIX 系统的稳定性, 本项目对中心 A,B,C 三个测试集合的不同严重程度下的 MAE 稳定性进行了评估 (图 3.11)。图 3.11 的折线图分别展示了 AI 模型在中心 A,B,C 上预测不同程度患者的 MAE (9 种模态配置类型, 带有 1 倍

标准误差)。这对于 AHI 预测的不同容限范围提出了要求，对于 OSA 严重程度增加导致的截断值非线性变化，AI 模型对不同程度患者 MAE 合理的稳定分配能力是提升分级性能的关键。本项目将 MAE 值根据所在严重程度 AHI 的上下界进行了缩放，结果表明了 AIX 系统在类型 I,II,V 都表现了良好的稳定性(折线趋势趋于平缓)，为通过交互的方式实现专家级诊断准确率提供了保障。

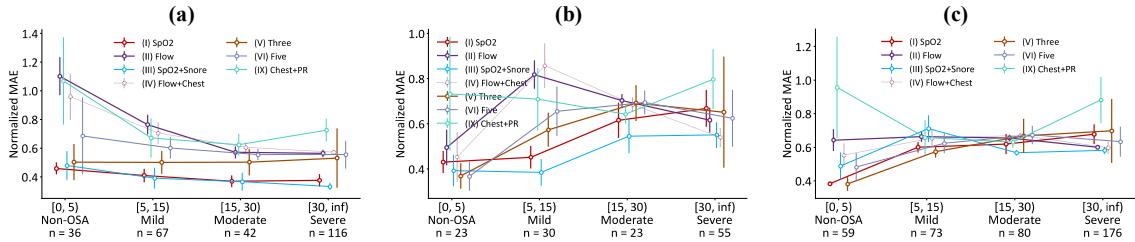


图 3.11 不同中心 AIX 在不同严重程度下的 AHI 评估稳定性

## (2) AIX 系统可视化和可解释性

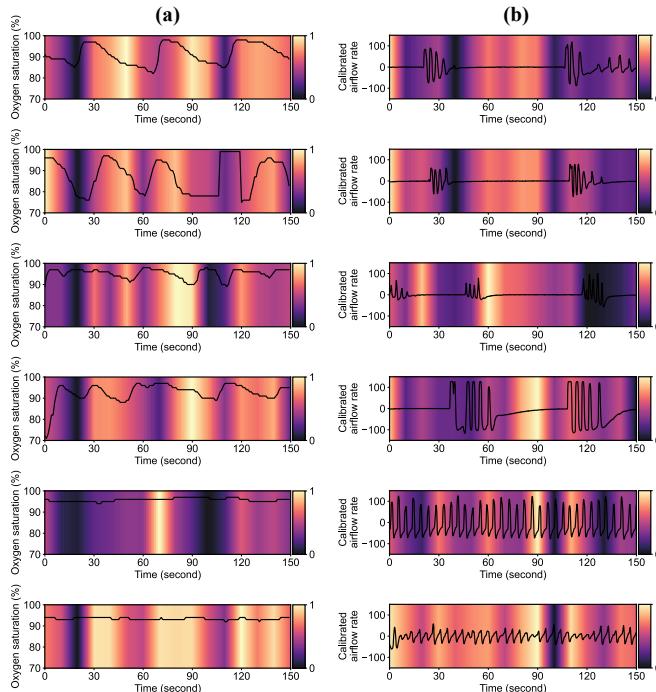


图 3.12 模型解释性注意力可视化

本项目通过在测试集上可视化注意力矩阵来提升模型对呼吸事件模式的解释能力。对从 AI 模型输出得到的注意力矩阵沿着特征图维度进行平均，并通过线性插值对其进行重新采样，使其与信号段的长度相匹配，以热力图的形式直观地展示了模型关注的区域(图 3.12)，模型关注的区域呈现出更亮的标记。在可视化之后，本项目观察到呼吸事件段(1-4 行)和非呼吸事件段(5-6 行)之间的注意力区域存在明显的模式差异。本项目分别展示了与 OSA 特征最相关的两种模态信号

(SpO<sub>2</sub> 和 Flow)，分别由图 3.12 (a) 和图 3.12 (b) 表示。在 SpO<sub>2</sub> 模态信号中，异常呼吸事件模型关注与血氧下降相关的区域，表明异常或非典型的血氧饱和度模式，暗示可能存在呼吸问题或疾病。相反，正常信号模型中感兴趣的区域反映了正常健康条件下观察到的典型血氧饱和度模式，通常是稳定的区域，没有明显的下降或变化。这些区域在规律性、稳定性或典型的血氧变化模式表现出与异常呼吸事件特征的不同。在 Flow 模态信号中，异常呼吸事件模型强调了受限或完全消失的气流区域，可能表明气道阻塞、停滞或不规则呼吸模式。相反，正常信号模型中感兴趣的区域代表正常健康呼吸模式中连续和规律的气流。本项目呈现的与模型注意力相关的可视化结果验证了模型检测到的模式与当前时间段内临床专家评估的强一致性。这种验证有助于提升模型的解释能力，让医生了解 AI 模型推理逻辑，增强决策信心。

此外，本项目基于 t-SNE 可视化了基于 9 种模态配置类型 (I-IX) 的 AI 模型学习的段级事件特征 (图 3.13)，顶部子图展示了原始信号特征，表现了不可区分特性。随着模型的特征学习进程，本项目观察到除了 VI 和 VIII 之外的类型都有较为明显的类边界，提供了有些模态类型无法区分两类事件导致性能较差的证据。

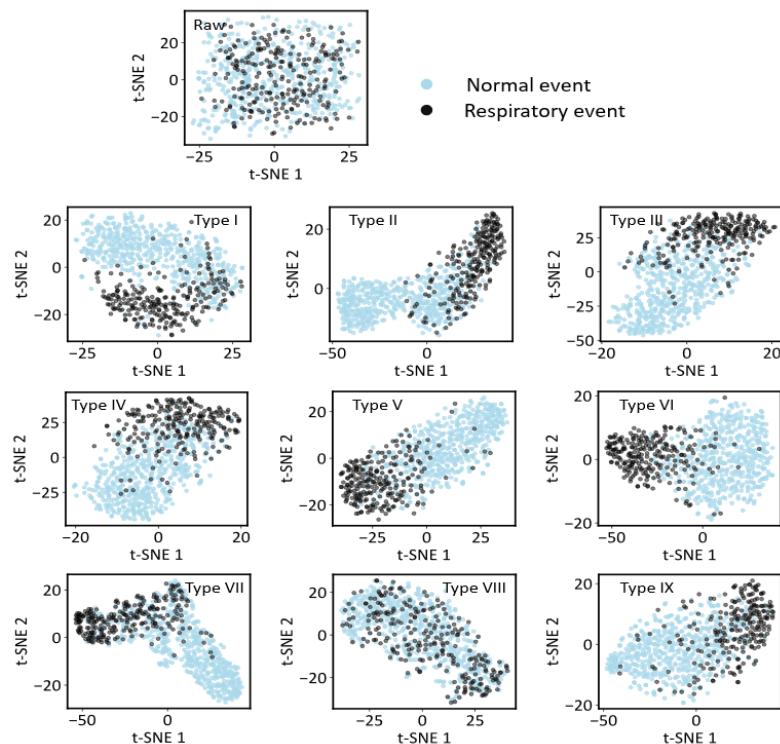


图 3.13 模型学习特征 t-SNE 可视化

### (3) AIX 系统可交互性验证

此外，本项目还探讨了 AI 模型交互的可行性。图 3.14 展示了模型正确 (TN 和

TP) 和错误 (FN 和 FP) 段级事件预测的聚类特征，分别来自不同的中心 A,B,C 的单个代表性受试患者的整夜段级事件。本项目观察到，错误的预测主要集中在两个聚类之间的中间区域，这解释了 AI 模型预测失败的原因。本项目对中心 A,B,C 的分段级事件预测结果进行显著性分析 (Two-tailed Mann-Whitney U-tests)，比较 AI 模型在正确和错误预测中预测两种事件类型 (正常事件和呼吸事件) 之间概率值差异的性能 (图 3.15)。结果表明，错误预测的概率绝对值差模式与正确预测的概率绝对值差模式存在显著差异，预测置信度较低 (即两种类型的段级事件概率接近)。这一特性为 AIX 呈现低预测置信度的样本供医生进行交互校正提供了依据。为了克服全自动预测的性能瓶颈，探索实现专家级诊断性能的交互条件，本项目从中心 A,B,C 中分别随机选择 50 例患者，组成一个包含 150 例患者的交互测试集。本项目定义了一个交互阈值 (两种事件类型的 AI 模型预测概率差异的绝对值，范围从 0.30 到 1.00)。本项目测试了 11 折交叉验证模型的三种交互模态类型 (类型 I, II 和 V)。假设医生可以根据视觉呼吸事件特征进行完全纠正，本项目评估了 OSA 四类分级性能 (图 3.15b,c,e)，以及不同交互阈值下每位患者的平均交互时间 (假设单次交互持续 0.5 秒) (图 3.15d)。基于系统交互的可行性，将在以下部分对 AIX 进行多中心临床交互验证。

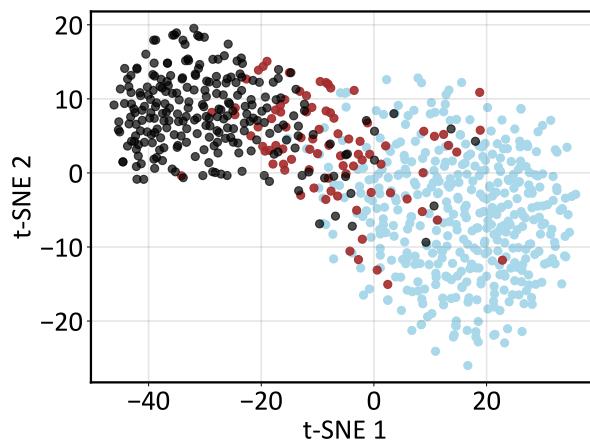


图 3.14 模型错误预测聚类散点图

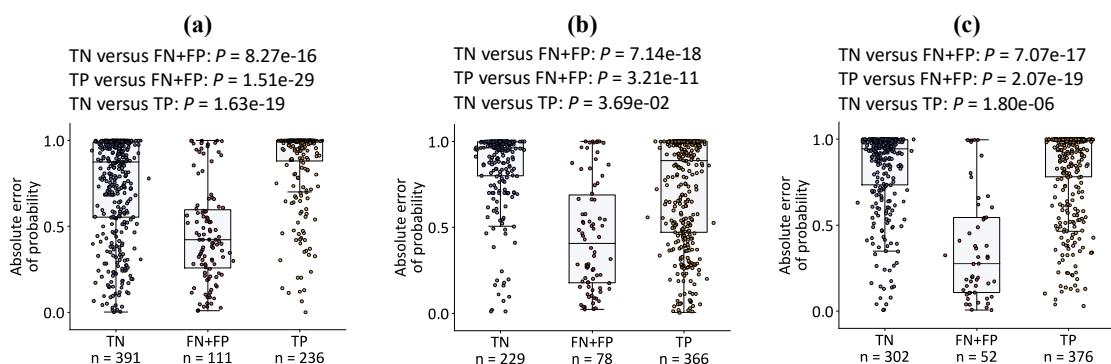


图 3.15 模型可交互性验证

#### (4) AIX 系统临床验证

本项目进一步测试了 AIX 的临床交互能力，搜集了来自中心 A,B,C 的各 50 例患者，进行多中心交互验证，AIX 系统被安装至临床医生的个人笔记本电脑上，并提供了一键导出呼吸事件报告的功能（包含模型对整晚数据的预测概览，异常呼吸时间占比，AHI 指数，连续异常事件个数等），为医生提供基于智能化技术的快速报告出具。Fig. 3.16 (a) 展示了本项目将 AIX 系统安装在临床中供医生操作的示例，本项目为 AIX 系统设计了医生友好的交互界面，通过选取可访问的模态组合类型，以调用不同的 AI 模型。可视化区域包括了原始信号回溯，通过滑块和加减操作及时定位事件索引，注意力图展示了模型关注区域，这增强了 AI 模型预测原理的透明度。每个段级事件对应的正常异常概率在右上角被展示，下方是对应患者的 AHI，整晚数据的 SpO<sub>2</sub> 变化（如模态配置包含该模态信号）和段级事件预测（黄色为正常时间，紫色为异常时间，以 30 秒粒度显示）总览被展示，下方是通过‘YES’和‘NO’来修改当前段级事件是否为异常。最后，提供了报告导出用于一键报告生成。上述为 AIX V1.0 版本，然而在医生进行实际交互操作时，结合了医生快速处理的需求，本项目将 AIX V1.0 升级为 AIX V2.0，具体的改动为左上角原始信号片段被拉长以提供更详尽细节，上方显示了当前事件的 AI 预测结果（正常或异常），同时下方加入了交互样本索引搜索功能，通过搜索满足交互阈值条件的样本，以便医生更快速的矫正 AI 模型预测（图 3.17）。

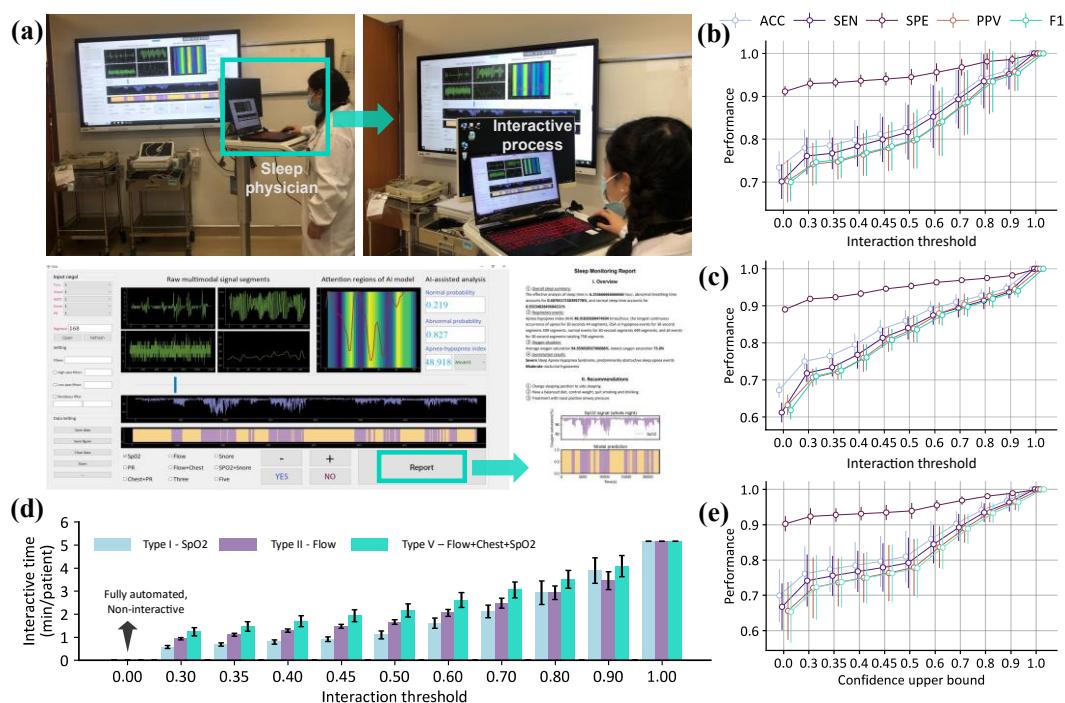


图 3.16 AIX 临床验证及交互性能时间对比



图 3.17 AIX V2.0 界面

图 3.17 和表 3.4 展示了在来自三个中心的 150 名患者的前瞻性实验中验证的结果（基于模态类型 V，其为临床医生最常使用的组合类型且能提供可靠的睡眠呼吸分析），对 OSA 四类严重程度分级展示了准确率，平均 F1 值以及单个患者平均处理时间（分钟）。可以看出随着 AIX 系统交互阈值的提升，其性能越来越接近专家级诊断，在交互阈值为 0.9 时，为临床医生缩短了约 6.9 倍处理时间，当阈值为 1.0 时，缩短了约 3.7 倍处理时间。这一结果表明了 AIX 系统具有极大的临床效率提升价值，且能实现专家级性能，出具可靠，快速的智能化报告。

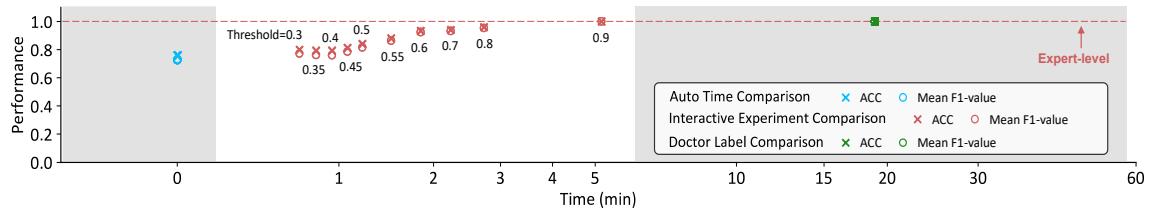


图 3.18 AIX 系统前瞻性性能评估

表 3.4 AIX 专家级诊断性能临床前瞻性评估

评估指标	临床专家	AIX 系统全自动预测	交互阈值									
			0.3	0.35	0.4	0.45	0.5	0.6	0.7	0.8	0.9	1.0
准确率	1	0.760	0.80	0.793	0.793	0.813	0.84	0.88	0.933	0.94	0.96	1
平均 F1 值	1	0.727	0.772	0.762	0.761	0.786	0.816	0.864	0.926	0.933	0.954	1
平均时间 (分钟)	18.9	0	0.688	0.812	0.941	1.074	1.211	1.501	1.839	2.229	2.723	5.165

### 3.3 技术可行性分析及创新说明

#### 3.3.1 技术可行性分析

有效诊断和管理 OSA 需要高效、准确且经济可接受的技术手段。但目前，OSA 的诊断和管理面临多重挑战。传统的 PSG 设备昂贵且设备数稀少，且需要专门的操作培训，导致高昂的检测费用和对资源的高度依赖。同时，预约睡眠监测的排队时间通常较长，很多患者需要等待数周甚至数月才能接受检查，这大大延误了诊断和治疗的进程。此外，目前鲜有 AI 辅助的自动化算法能在 OSA 诊断中发挥显著作用。现有的全自动算法泛化性能具有瓶颈，无法提供足够精确和可靠的诊断结果。

而本项目介绍的 AIX 系统设计并集成了低成本信号采集硬件设备以及先进的 AI 算法，旨在通过快速的人机交互提供专家级诊断准确性，并提供临床可解释性。以下从各个方面详细分析该系统的技术可行性。

##### (1) 多模态生理信号采集可行性分析

在现代医学诊断中，多模态生理信号的采集日益成为研究和应用的热点。通过融合多种生理信号，不仅可以提高诊断的准确性和全面性，还能提供更加丰富的病情信息，为临床决策提供有力支持。本项目所设计的硬件系统，主要包括 ECG 信号和 PPG 信号的采集，通过这两种信号的协同分析，可以显著提升对 OSA 的监测和诊断效果。

**ECG 信号采集：**ECG 信号采集部分使用了 AD8232 集成电路，其具备高精度的信号放大与滤波功能。具体设计包括：1) 信号预处理，通过电极贴片采集的心电信号，首先输入到 AD8232 进行初级放大和高通滤波。高通滤波器的设计采用双极点配置，截止频率设定为 0.05Hz，用于去除基线漂移和低频噪声。2) 信号放大与滤波，初级放大后的信号通过 Sallen-Key 低通滤波器进一步滤除高频噪声，随后进行二级放大，整体放大倍数达到 1000 倍。低通滤波器的截止频率设定为 45Hz，以有效去除肌电干扰和运动伪迹。这种设计确保了心电信号的高质量采集和处理，为后续的数字化和分析提供了可靠的数据基础。

**PPG 信号采集：**PPG 信号的采集采用了 MAX30102 传感器模块，该模块集成了红外和近红外光源、光电探测器以及高性能的模数转换器（ADC）。其主要特点包括：1) 高分辨率采样，MAX30102 内部的 ADC 具备 18 位分辨率，支持高达 10.24MHz 的采样频率，并可编程数据输出率范围为 50 至 3200sps，确保了信号采集的高精度。2) 环境光消除：内置的环境光消除功能，有效减少外部光源对信号的干扰，提升了信号的纯净度。3) 模块化设计：传感器模块通过 I2C 接口与主控芯片进行通信，红光及近红外 LED 的驱动和供电引脚经过优化设计，确保了系统的稳定性和可靠性。通过上述设计，脉搏波信号的采集达到了高精度和高稳定性的

要求，能够实时反映受试者的脉搏变化情况。

**多模态信号的应用与优势：**多模态生理信号采集的一个重要优势在于能够提供更为全面的生理状态信息。优势包括：1) 综合评估，ECG 信号和 PPG 信号的结合，可以从不同角度反映心脏和循环系统的状态。例如，ECG 信号能够揭示心脏的电活动情况，而 PPG 信号则可以反映血液流动的动态变化。2) 提高诊断准确性，在 OSA 的监测中，单一信号可能受限于某些特定因素的影响，而多模态信号的融合使用，可以提供更为全面和准确的诊断依据。例如，ECG 信号的异常变化可能与呼吸暂停事件有关，而脉搏波信号的变化则可以进一步确认这一事件的发生。3) 数据冗余与可靠性，多模态信号的采集增加了数据的冗余性，即使某一信号在特定情况下出现缺失或干扰，其他信号仍然可以提供有效的信息，保证了整体监测系统的可靠性。因此，基于 ECG 和脉搏波信号的多模态生理信号采集方案，不仅具有可行性，而且在提升 OSA 监测和诊断的准确性、全面性以及系统的可靠性方面，展示出了显著的优势。这为 OSA 的早期诊断和干预提供了强有力的技术支持。

## (2) 算法和系统可行性分析

本项目提出了一种基于 AI 的轻量级进化注意力模型，用于准确、舒适和高效地诊断 OSA。并且设计了 AIX 系统以集成模型并提供智能的交互可解释界面。

**数据集构建：**数据集的构建是模型训练和评估的基础。本项目从 2020 年到 2024 年期间收集了来自三个不同中心 A,B,C 的共计 2191 名疑似 OSA 患者的夜间多模态记录，共计超过 10,000 小时的数据。这些数据包括 Flow、Chest 和 SpO2 等多种模态信号。为了训练 AI 模型，需要将睡眠数据进行标注。本项目从睡眠阶段中提取 30 秒的注释，去除清醒数据，并进一步消除干扰信号。然后将这些时间戳注释转换为 30 秒的呼吸事件标签。呼吸事件的段级检测结果被转换为呼吸暂停低通气指数 (AHI)，并进一步获得患者级 OSA 的四级分类。

数据集的优势在于其大规模和多样性，涵盖了不同类型的呼吸事件和正常呼吸周期，提供了一个全面的样本库，以支持模型的训练和评估。此外，本项目采用了 11 折交叉验证的方法，以确保模型在不同患者数据上的泛化能力，增强了数据集的代表性和模型的稳健性。

**灵活模态简化配置可行性：**在资源受限的环境中，传统的多模态信号采集设备可能难以普及，因此利用 AI 技术挖掘在信息受限条件下的高级表征，实现高性能的 OSA 诊断，具有重要的现实意义。为了验证简化配置信号在不同场景下的适用性，本项目在不同模态组合上进行了广泛测试，结果显示即使在简化配置下，模型仍能保持较高的诊断性能。这为在广泛的 OSA 筛查中采用更简洁的信息选择方法提供了科学依据。通过这些分析，可以看出在不同资源条件下，模型都能够灵活地调整信号采集配置，从而实现高效、准确的 OSA 诊断。这种灵活性不仅降低了设

备成本和技术门槛，还提高了诊断的普及性和可操作性，为大规模 OSA 筛查提供了可行的解决方案。

**AI 模型设计：**本项目提出了一种新型的 LEAF-Net 模型，该模型通过从各种模态信号中可靠地提取局部-全局生理特征模式，为不同场景下 OSA 的准确诊断提供了前提条件，并提供了一种 OSA 严重程度分级的自动化方法。本项目评估了 9 种不同模态选择在段级呼吸事件检测中的性能。这些模态包括单一的 SpO<sub>2</sub> 信号、双通道的 SpO<sub>2</sub> 和打鼾信号、以及常用的 Flow+Chest+SpO<sub>2</sub> 组合。研究发现，单一的 SpO<sub>2</sub> 信号、SpO<sub>2</sub> 和打鼾信号的双通道组合，以及 Flow+Chest+SpO<sub>2</sub> 组合在二元事件分类任务中表现最佳。此外，单一的 Flow 信号和 Chest+PR 组合也表现出有效的二元事件分割结果。

为了评估模型的稳健性，本项目将测试集划分为 11 个子集，并使用每个子集上验证集损失最低的模型进行评估。模型性能通过 ROC 曲线和 PRC 曲线，以及 ACC、F1、AUROC、AUPRC 等四个指标进行评估。结果表明，简化信息融合的复杂性仍然可以有效区分二元事件，这为在广泛的 OSA 筛查中采用更简洁的信息选择方法提供了参考。

**AI 模型交互可解释性：**本项目基于 AI 驱动的生理增强模型，搭建了一个交互式和可解释的临床诊断系统。该系统不仅能自动处理整夜多模态数据，还能一键生成 OSA 诊断相关的报告，并促进大规模患者数据集的快速和可解释处理分析。同时，为了使临床医生能够更好地理解和使用模型，本项目开发了一个临床友好的交互式用户界面，该界面以 AI 驱动的方式呈现。这种交互性不仅减轻了医生的负担，还缩短了睡眠报告的发布时间，并具有并行处理大量数据的能力。

### 3.3.2 创新说明

这篇工作的创新点主要集中在以下几个方面：

#### (1) 大规模多中心临床数据集的应用

研究收集了 2191 名患者的整夜多模态记录，总计超过 10,000 小时的多模态信号，用于训练和测试提出的 AI 模型系统。在临床环境中，该系统展示了其作为临床工具的价值，显著提高了 OSA 诊断中的监测舒适度和效率。

#### (2) AI 驱动的轻量级进化注意力模型

本项目提出了一种轻量级注意力模型，用于实现更舒适的睡眠监测和高效、透明的 OSA 评分。模型通过从多种模态信号中可靠地提取局部和全局的生理特征模式，为不同场景下准确诊断 OSA 提供了前提条件，并提供了一种 OSA 分级的自动化方法。

#### (3) 多模态信号灵活简化配置探索

通过跨模态配置组设计，研究了不同信号组合对 AI 模型在 OSA 诊断准确性上的影响，得出了模式简化的结论，为更舒适的监测配置提供了实验支持。这种模式简化不仅提高了患者的监测舒适度，还保持了诊断的高准确性。

#### **(4) 针对不同监测场景的硬件设计**

本项目为不同监测场景设计了两套硬件采集设备。第一套用于较专业场景，模态丰富，制造成本较低，能够在关键诊断中提供高精度数据，其高性能和准确性对临床诊断非常有价值。第二套用于日常场景，模态较少，设计简洁，制造成本低，适用于可穿戴、动态、便携的长程监护。也有利于大规模推广和普及。两套设备分别针对不同的监测场景进行优化，使得系统能够灵活适应从日常健康监测到专业医疗诊断的多种需求。

#### **(5) 交互可解释的专家级 OSA 诊断系统 AIX**

本项目开发了一个交互可解释的专家级 OSA 诊断系统，提供了一个临床友好的交互用户界面，医生可通过人机协作的方式达到专家级诊断准确率。该系统能够自动批量处理整夜的多模态数据，生成 OSA 诊断相关的报告，并促进对大量患者数据的快速和可解释处理。这大大减轻了医生的负担，提高了诊断效率。

#### **(6) AIX 系统的临床验证及优化**

本项目对 AIX 系统进行了全面的临床验证和优化。通过在多中心临床环境中的实际应用，本项目采集了 150 例多中心真实患者的数据，并根据临床反馈不断改进系统性能。这一过程不仅验证了 AIX 系统在不同临床场景中的有效性，还通过优化算法和硬件配置，提升了系统的稳定性和诊断准确性。临床验证显示，AIX 系统能够显著提高 OSA 诊断效率，并增强医生的决策支持能力，从而为临床应用提供了有力的实证支持。

## 4 总结

### 4.1 作品展示

本项目设计并评估了一个 AI 驱动的交互可解释性阻塞性睡眠呼吸暂停专家级 OSA 诊断系统 AIX。通过在医院给睡眠医生个人电脑安装打包好的软件格式，通过投屏展示了用户操作友好的上位机操作（图 4.1）。AIX 系统可以自动批处理整夜多模态数据，一键生成 OSA 智能化诊断报告，有助于快速解读大规模患者数据。上位机界面设计友好，医生可以轻松地进行交互操作。生成的报告通过 AI 驱动，极大地减轻了医生的工作负担，缩短了睡眠报告发布的时间，同时具备并行处理海量数据的能力。这一展示了在医疗 5.0 时代，AI 技术在提高患者监测舒适度和医疗效率方面的关键作用。

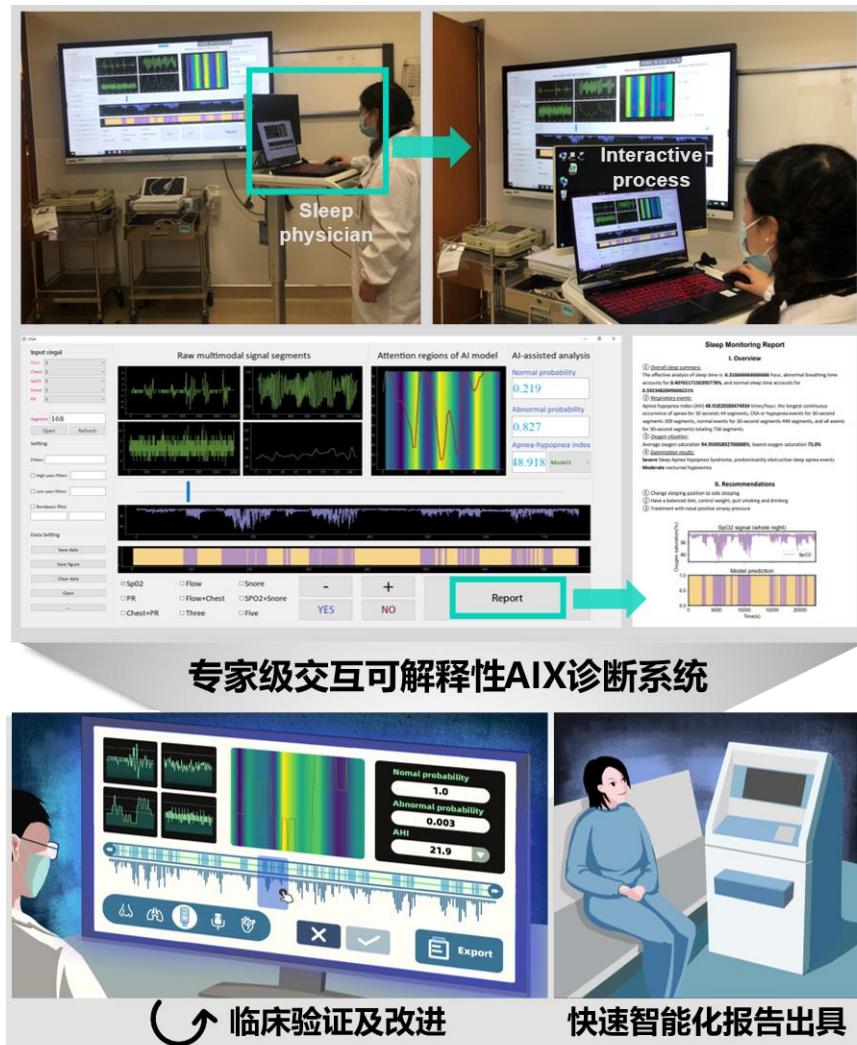


图 4.1 作品展示

## 4.2 展望

随着 AI 技术的飞速发展，医疗领域迎来了前所未有的变革。本项目设计了一种基于 AI 的轻量化进化注意力模型和一个交互可解释的专家级 OSA 诊断系统。这些创新展示了 AI 在提高诊断效率和患者舒适度方面的巨大潜力。然而，要将这些技术从小规模应用阶段转变为临床实践的广泛应用中，还有许多工作需要完成。以下是一些未来研究和应用的展望：

### (1) 数据集的扩展与多样化

目前的研究基于 2191 名患者的多中心多模态数据集。尽管这是一个相对较大的数据集，但为了进一步验证和提升模型的泛化性和鲁棒性，需要收集更多样化的患者数据。这包括不同年龄段、性别、种族和健康状况的患者数据。此外，跨地域的数据采集也非常重要，因不同的地理和文化背景会导致睡眠行为改变和 OSA 表征差异。

### (2) 模型个性化

每个受试者的生理特征和睡眠模式有所不同，个性化模型的开发将进一步提升诊断准确性。未来的研究可以考虑以下方面：

**个体特征建模：**根据患者的个体特征（如年龄、性别、体重指数等）微调模型参数，使其能够更好地适应不同患者的具体情况。

**自适应学习：**引入自适应学习技术，使模型能够在不断接收新数据的过程中自动调整和优化，从而提高对个体差异的适应能力。

**长时间跟踪：**通过长期跟踪患者的睡眠数据，建立个性化的健康档案，为每个患者量身定制诊断和治疗方案。

### (3) AIX 系统的大规模验证与应用

本项目未来的研究计划在更多的医疗机构和更广泛的患者群体中进行大规模临床试验，验证 AIX 系统的稳定性和有效性，并与现有医疗信息系统集成，通过优化用户界面和操作流程继续提升系统使用舒适性，并不断收集临床医生的反馈和运行监测数据以改进系统性能，确保其在实际应用中的高效和可靠。