

基于多时间粒度深监督的SpO₂睡眠呼吸暂停检测方法

张德军, 杨其宇

广东工业大学自动化学院, 广东 广州

收稿日期: 2026年4月22日; 录用日期: 2026年5月19日; 发布日期: 2026年5月27日

摘要

血氧饱和度(SpO₂)信号能够反映睡眠过程中由异常呼吸事件引起的血氧下降与恢复过程, 在睡眠呼吸暂停检测中具有重要应用价值。与脑电等生理信号相比, SpO₂获取简便且适合长期连续监测, 但现有方法多基于单一时间粒度分析, 难以兼顾局部细节与长程时序结构。针对上述局限, 文章提出一种基于多时间粒度深监督的SpO₂睡眠呼吸暂停检测方法。该方法以1小时、1 Hz采样的SpO₂序列为输入, 构建一维U-Net编码器-解码器结构, 并在1 s、5 s、15 s、30 s和60 s五个时间尺度上引入监督分支, 通过多尺度标签构建与联合优化增强模型对跨尺度异常模式的表征能力。实验结果表明, 所提方法在SpO₂单模态任务上具有良好稳定性, 五折交叉验证平均准确率为79.96%, 平均F1-Score为77.39%。这证明多时间粒度建模与深监督联合优化可提升模型对呼吸暂停事件的识别能力, 为低侵入式睡眠呼吸暂停筛查提供了一种有效方法。

关键词

血氧饱和度, 睡眠呼吸暂停检测, 多时间粒度建模, 深监督, 一维U-Net

SpO₂ Sleep Apnea Detection Method Based on Multi-Time Granularity Deep Supervision

Dejun Zhang, Qiyu Yang

School of Automation, Guangdong University of Technology, Guangzhou Guangdong

Received: April 22, 2026; accepted: May 19, 2026; published: May 27, 2026

Abstract

Blood oxygen saturation (SpO₂) signals can reflect the desaturation and recovery processes caused

by abnormal respiratory events during sleep, and thus have important application value in sleep apnea detection. Compared with physiological signals such as EEG, SpO₂ is easy to acquire and suitable for long-term continuous monitoring; however, existing methods are mostly based on a single temporal resolution, making it difficult to capture both local details and long-range temporal structures. To address this limitation, this paper proposes a multi-temporal granularity deep supervision-based method for SpO₂ sleep apnea detection. The method takes a 1-hour SpO₂ sequence sampled at 1 Hz as input, constructs a one-dimensional U-Net encoder-decoder architecture, and introduces supervision branches at five temporal scales (1 s, 5 s, 15 s, 30 s, and 60 s). A multi-scale label construction and joint optimization strategy is adopted to enhance the model's ability to represent cross-scale abnormal patterns. Experimental results show that the proposed method achieves good stability in the single-modality SpO₂ task, with an average accuracy of 79.96% and an average F1-Score of 77.39% under five-fold cross-validation. These findings verify that the combination of multi-temporal modeling and deep supervision improves the detection of apnea events, providing an effective approach for low-invasive sleep apnea screening.

Keywords

SpO₂, Sleep Apnea Detection, Multi-Temporal Granularity Modeling, Deep Supervision, 1D U-Net

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

睡眠呼吸暂停相关事件会导致夜间血氧饱和度下降、恢复延迟及波动模式异常, 因此血氧饱和度 (SpO₂) 信号在睡眠呼吸异常检测中具有重要应用价值[1]。与脑电、气流或胸腹运动等生理信号相比, SpO₂ 采集方式简便、设备负担较低, 更适用于家庭场景下的长期连续监测。然而, SpO₂ 对呼吸事件的响应通常具有一定滞后性和时间扩散特征[2], 这使得传统基于单一时间粒度的检测方法难以充分刻画事件的层次化时序模式。

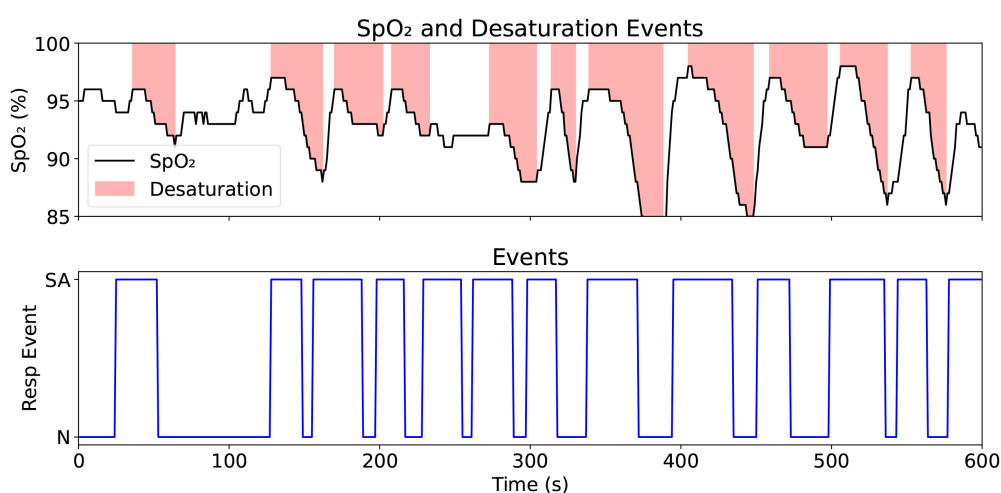


Figure 1. Schematic illustration of the temporal alignment among SpO₂ variations, oxygen desaturation events, and respiratory events

图 1. SpO₂ 变化及氧减事件与呼吸事件的时间对齐示意图

从生理表现来看, 呼吸暂停相关异常既包含逐秒级的局部低氧下降边缘、恢复过程和短时波动, 也包含十秒级至分钟级的持续低氧、周期性起伏增强等现象, 如图 1 所示。因此, 仅依赖单一时间粒度建模容易忽略局部细节或长程上下文中的部分信息, 从而限制模型对异常模式的完整表征能力。近期研究进一步表明, 从 1 s 到数百秒不同时间尺度提取 SpO_2 特征有助于提升睡眠呼吸暂停检测和严重程度分类性能, 尤其是部分非线性特征在较粗时间尺度下仍保持较强判别能力, 这说明血氧异常模式具有明显的跨尺度时间结构[2]。

针对上述问题, 本文提出一种基于多时间粒度深监督的 SpO_2 睡眠呼吸暂停检测方法。该方法以一维 U-Net 为主体结构, 在不同分辨率的解码层上引入多粒度分类头, 并构建与之对应的 1 s、5 s、15 s、30 s 和 60 s 标签体系, 通过联合优化提升模型的跨粒度表征能力。相较于传统仅在最终输出层进行检测的方式, 该方法能够使中间层特征直接接受与其语义层次一致的监督, 从而增强对局部异常与长程持续结构的协同建模能力。

本文的主要贡献如下:

- 1) 提出一种面向多时间粒度事件表征的 SpO_2 睡眠呼吸暂停检测框架, 在统一网络中同时建模细粒度异常波动与粗粒度持续模式。
- 2) 设计多粒度标签构建策略, 使不同时间粒度的监督信号与一维 U-Net 各层特征分辨率自然对齐, 增强模型训练过程中的语义一致性。
- 3) 通过多粒度深监督与联合优化策略, 提高不同层级特征的判别能力, 并在单模态 SpO_2 条件下获得稳定的检测性能。

2. 相关工作

2.1. 基于人工特征的 SpO_2 呼吸暂停检测方法

早期基于 SpO_2 的睡眠呼吸暂停检测研究主要依赖手工特征提取与传统机器学习分类器相结合的方法。Rolón 等人[3]从血氧信号中提取统计特征和形态学特征, 如均值、方差以及脱饱和事件的幅度和持续时间, 并结合支持向量机、随机森林和神经网络等分类器实现呼吸暂停事件检测。Liu 等人[4]则通过互信息分析与特征选择, 结合集成学习模型完成阻塞性睡眠呼吸暂停筛查。Sharma 等人[5]利用小波变换提取时频特征, 并结合 BaggedTree 分类器以增强对血氧波动模式的识别能力。总体而言, 该类方法通常围绕血氧下降幅度、持续时间、恢复过程、氧减指数及局部波形统计量构建判别特征, 并借助阈值规则、逻辑回归、支持向量机或其他浅层分类器完成检测任务。然而, 这类方法高度依赖经验设计, 特征表达能力有限, 对信号噪声、个体差异及标注标准较为敏感。当呼吸事件边界模糊、低氧响应较弱或存在明显时滞时, 预定义特征往往难以稳定刻画异常模式, 从而限制了模型的泛化能力。

2.2. 基于深度学习的 SpO_2 呼吸暂停检测方法

随着深度学习在生理信号分析中的广泛应用, 越来越多研究开始采用数据驱动方式对原始或预处理后的 SpO_2 序列进行端到端建模。John 等人[6]提出了基于 1D-CNN 的 SomnNET 模型, 能够直接处理原始血氧信号, 实现了对血氧信号的高分辨率检测。Deviaene 等人[7]提出了一种融合全夜特征的 BiLSTM-CNN 模型, 结合了 CNN 的特征提取能力和 BiLSTM 的时间序列建模能力, 用于血氧信号的 OSA 筛查。Almarshad 等人[8]探索了基于 Transformer 的神经网络结构在血氧信号诊断呼吸暂停中的应用, 以利用其强大的自注意力机制捕捉长程依赖关系, 并比较了 1 s 至 360 s 不同时间分辨率下的诊断性能, 表明时间粒度设计会显著影响模型表现。Hou 等人[9]提出了 OSASformer, 通过 Transformer 双分支网络融合多源表示, 在可穿戴场景下实现了较好的 OSA 筛查性能, 表明自注意力机制在长程时序依赖建模方面具有应

用潜力。

尽管如此, 现有深度学习方法在 SpO_2 事件检测任务中仍存在一定局限。一方面, 许多研究主要围绕单一时间尺度建立监督目标[10], 检测结果往往集中于某一固定粒度, 难以同时兼顾逐秒级边界变化与十秒级至分钟级持续模式。另一方面, 部分方法仅在最终输出层施加监督, 中间层特征缺乏与其语义层次相一致的直接约束, 导致网络虽然具备较深层级结构, 却未能充分发挥多层表示在跨尺度事件建模中的能力。

3. 方法

3.1. 问题定义

设输入血氧序列为 $\mathbf{x} \in \mathbb{R}^{B \times 1 \times T}$, 其中, B 表示批大小, $T = 3600$ 表示 1 小时血氧信号在 1 Hz 采样率下的序列长度。与输入信号对应的原始逐秒呼吸事件标注表示为 $\mathbf{y}^{1s} \in \{0, 1\}^{B \times T}$, 其中, 标签 0 表示该时刻不存在呼吸暂停事件, 标签 1 表示该时刻存在呼吸暂停事件。为了刻画呼吸事件在不同时间尺度上的表现形式, 本文在逐秒标签基础上进一步构建 5 s、15 s、30 s 和 60 s 粒度的事件标签, 分别记为 $\mathbf{y}^{5s}, \mathbf{y}^{15s}, \mathbf{y}^{30s}, \mathbf{y}^{60s}$, 相应地, 模型输出多个时间粒度下的预测结果: $\hat{\mathbf{y}}^{1s}, \hat{\mathbf{y}}^{5s}, \hat{\mathbf{y}}^{15s}, \hat{\mathbf{y}}^{30s}, \hat{\mathbf{y}}^{60s}$ 。

3.2. 多时间粒度标签构建

为使模型在不同时间尺度下均获得明确监督信息, 本文在逐秒标签基础上构建多粒度事件标注。对于任意目标时间粒度 $g \in \{5, 15, 30, 60\}$, 将长度为 T 的逐秒标签划分为若干个长度为 g 的非重叠时间窗口。设第 k 个窗口内的事件累计值为

$$s_k^{(g)} = \sum_{t=(k-1)g+1}^{kg} y_t^{1s}$$

当窗口内事件点数不小于预设阈值 τ_g 时, 将该窗口标记为事件, 否则标记为正常, 即

$$y_k^{(g)} = \begin{cases} 1, & s_k^{(g)} \geq \tau_g, \\ 0, & s_k^{(g)} < \tau_g. \end{cases}$$

考虑到不同时间尺度下事件持续特征与标签稀疏程度的差异, 本文采用表 1 所示的标签映射规则:

Table 1. Rules for multi-temporal-granularity label construction

表 1. 多时间粒度标签构建规则

时间粒度	窗口长度 g	判定阈值 τ
1 s	1	1
5 s	5	1
15 s	15	3
30 s	30	5
60 s	60	10

该标签构建方式将原始逐秒标注映射为多个尺度下的片段级监督信息, 使模型能够同时学习短时异常变化和较长时间范围内的事件持续模式。

3.3. 多粒度 SpO_2 呼吸暂停检测网络

如图 2 所示, 本文采用一维 U-Net 作为 SpO_2 呼吸暂停检测的基础网络。U-Net 通过编码器逐层提取高层语义信息, 并借助解码器逐步恢复时间分辨率, 同时利用跳跃连接融合浅层局部细节特征与深层语

义特征, 适用于时序检测任务中局部模式与全局上下文并重的建模需求。模型具体细节如表 2 所示。

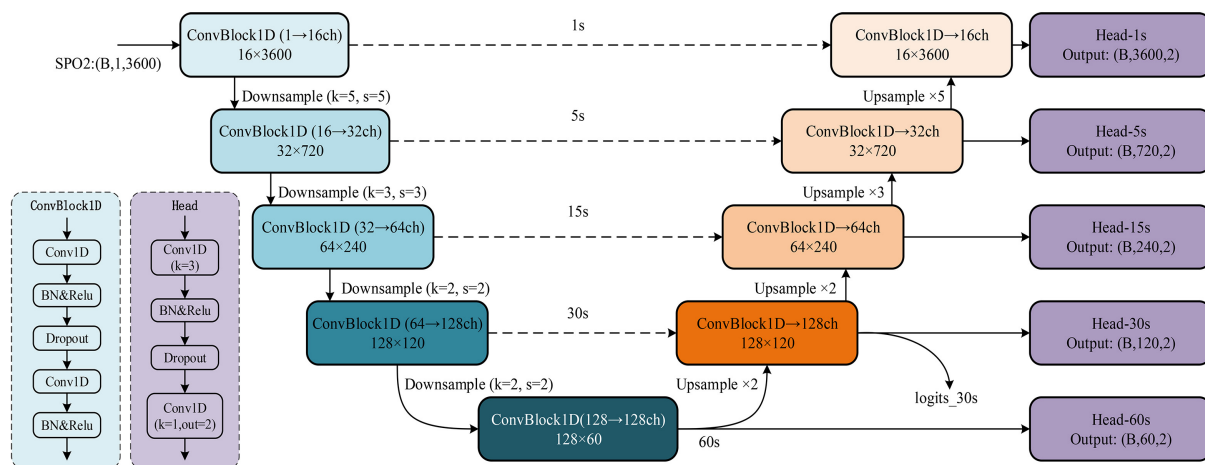


Figure 2. Framework of the proposed multi-temporal-granularity SpO₂ sleep apnea detection method

图 2. 多时间粒度血氧呼吸暂停检测框架

Table 2. Architecture of the 1D U-Net network

表 2. 一维 U-Net 网络结构

模块	操作	输出尺寸	说明
输入层	原始血氧序列	$B \times 1 \times 3600$	1 小时 SpO ₂ , 采样率 1 Hz
编码器 1	ConvBlock	$B \times 32 \times 3600$	提取逐秒局部特征
下采样 1	Downsample ($\times 5$)	$B \times 64 \times 720$	对应 5 s 粒度
编码器 2	ConvBlock	$B \times 64 \times 720$	建模短时上下文
下采样 2	Downsample ($\times 3$)	$B \times 128 \times 240$	对应 15 s 粒度
编码器 3	ConvBlock	$B \times 128 \times 240$	提取中尺度时序模式
下采样 3	Downsample ($\times 2$)	$B \times 256 \times 120$	对应 30 s 粒度
编码器 4	ConvBlock	$B \times 256 \times 120$	建模片段级事件结构
下采样 4	Downsample ($\times 2$)	$B \times 512 \times 60$	对应 60 s 粒度
瓶颈层	ConvBlock	$B \times 512 \times 60$	高层语义特征提取
解码器 4	Upsample + Skip + Conv	$B \times 256 \times 120$	与编码器 4 跳跃连接
分类头(30 s)	1×1 Conv	$B \times 2 \times 120$	30 s 粒度事件分类
解码器 3	Upsample + Skip + Conv	$B \times 128 \times 240$	与编码器 3 跳跃连接
分类头(15 s)	1×1 Conv	$B \times 2 \times 240$	15 s 粒度事件分类
解码器 2	Upsample + Skip + Conv	$B \times 64 \times 720$	与编码器 2 跳跃连接
分类头(5 s)	1×1 Conv	$B \times 2 \times 720$	5 s 粒度事件分类
解码器 1	Upsample + Skip + Conv	$B \times 32 \times 3600$	与编码器 1 跳跃连接
分类头(1 s)	1×1 Conv	$B \times 2 \times 3600$	1 s 粒度事件分类
瓶颈输出头	1×1 Conv	$B \times 2 \times 60$	60 s 粒度事件分类

3.3.1. 编码器

编码器以原始血氧序列为输入, 通过多层一维卷积模块逐层提取层次化时序特征, 并在每一层通过下采样操作压缩时间维度、扩大感受野。设第 l 层编码特征表示为

$$E_l \in \mathbb{R}^{B \times C_l \times T_l},$$

其中, C_l 表示该层特征通道数, T_l 表示对应的时间长度。随着网络深度增加, 特征表示逐渐由局部波动模式向更高层次的语义表征过渡。

3.3.2. 解码器

解码器对深层特征进行逐层上采样, 以恢复时间分辨率, 并与编码器中相同尺度的特征进行跳跃连接。设第 l 层解码特征为

$$D_l \in \mathbb{R}^{B \times C_l' \times T_l'},$$

则其由上一层解码特征上采样后与对应编码层特征拼接, 再经卷积融合得到。通过该结构, 模型在恢复时序分辨率的同时能够充分利用浅层细节信息, 增强对事件边界和局部异常模式的刻画能力。

3.3.3. 多粒度分类头

为使不同层级的解码特征承担与其语义层次相匹配的检测任务, 本文在多个解码层后分别设置多粒度分类头。对于任意粒度 g , 分类头输出表示为

$$z^{(g)} = \text{Head}_g(D_l),$$

其中, $\text{Head}_g(\cdot)$ 表示对应粒度的分类映射操作。本文采用轻量级的 1×1 卷积将特征映射至二分类输出空间。各粒度输出的维度分别为

$$z^{1s} \in \mathbb{R}^{B \times 2 \times 3600}, z^{5s} \in \mathbb{R}^{B \times 2 \times 720}, z^{15s} \in \mathbb{R}^{B \times 2 \times 240}, z^{30s} \in \mathbb{R}^{B \times 2 \times 120}, z^{60s} \in \mathbb{R}^{B \times 2 \times 60}.$$

相较于仅在最终输出层进行预测的单尺度方法, 多粒度分类头能够使不同层次特征在训练过程中直接接受相应尺度的监督, 从而增强中间层表示的判别能力。

3.4. 多目标优化

由于模型在多个时间尺度下均输出对应的预测结果, 本文采用深监督策略对各粒度分类头进行联合优化。对于任意粒度 g , 定义其分类损失为

$$\mathcal{L}_g = \text{CE}(z^{(g)}, y^{(g)}),$$

其中, $\text{CE}(\cdot)$ 表示交叉熵损失函数, $z^{(g)}$ 为粒度 g 下的输出 logits, $y^{(g)}$ 为相应粒度的标签。在此基础上, 对所有粒度下的损失进行加权求和, 得到整体训练目标:

$$\mathcal{L}_{\text{spo2}} = \lambda_1 \mathcal{L}_{1s} + \lambda_5 \mathcal{L}_{5s} + \lambda_{15} \mathcal{L}_{15s} + \lambda_{30} \mathcal{L}_{30s} + \lambda_{60} \mathcal{L}_{60s}.$$

其中, λ_1 、 λ_5 、 λ_{15} 、 λ_{30} 和 λ_{60} 分别表示不同时间粒度损失项的权重系数。本文采用如表 3 所示的损失权重设置:

Table 3. Loss weight settings for different temporal granularities

表 3. 多粒度损失权重设置

时间粒度	损失权重
1 s	$\lambda_1 = 1.0$
5 s	$\lambda_5 = 0.8$
15 s	$\lambda_{15} = 0.6$
30 s	$\lambda_{30} = 0.5$
60 s	$\lambda_{60} = 0.4$

4. 实验设计

4.1. 数据集及预处理

本研究采用 Sleep Heart Health Study (SHHS) [11]数据集作为实验数据来源。考虑到实验规模控制与样本分布均衡性, 本文从该数据集的 v1 版本中筛选 500 例受试者样本用于实验, 并根据受试者的呼吸暂停低通气指数(AHI)进行统计分析。各类别样本的受试者人数、年龄、体质量指数(BMI)及性别分布如表 4 所示。

Table 4. Composition of the SHHS dataset

表 4. SHHS 数据集样本构成

类别	AHI 范围	受试者人数	年龄(岁)	BMI (kg/m ²)	性别(男/女)
无 OSA	AHI < 5	322	58.0 ± 8.7	26.8 ± 4.8	92/230
轻度 OSA	5 ≤ AHI < 15	76	59.9 ± 8.5	29.8 ± 4.8	41/35
中度 OSA	15 ≤ AHI < 30	55	62.9 ± 7.3	30.3 ± 5.2	32/23
重度 OSA	AHI ≥ 30	47	63.5 ± 7.6	32.9 ± 6.5	35/12
总体	-	500	59.3 ± 8.7	28.2 ± 5.4	198/302

为提高血氧(SpO₂)信号的质量并减少伪影干扰, 本文设计了一套预处理流程。在信号清洗阶段, 定义有效取值范围为 50%~100%, 将超出该范围的异常值及非有限值标记为无效, 并通过相邻采样点差分检测突变(阈值为 6%/s), 识别不合理跳变; 利用滑动窗口计算局部标准差, 检测长时间近似恒定的平坦信号段, 并一并标记为伪影。

对于缺失或异常片段, 采用分段处理策略: 当连续缺失长度不超过 30 秒时, 使用线性插值进行修复; 对于更长时间的缺失段, 则采用固定填充值(95%)进行填补。同时, 为增强信号稳定性, 先应用中值滤波去除孤立噪声, 再通过滑动平均滤波进行平滑处理。最终对信号进行范围裁剪, 以确保其落在生理合理区间内。

4.2. 评估指标

考虑到睡眠呼吸暂停检测任务中正常片段与异常事件片段通常存在明显的类别不平衡, 仅采用准确率难以全面反映模型对呼吸暂停事件的识别能力。因此, 本文从分类正确性、正类识别能力及整体判别性能等方面对模型进行综合评估, 选取准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F1 值(F1-Score)作为评价指标。

设真正例、假正例、真负例和假负例分别为 TP 、 FP 、 TN 和 FN , 则各指标定义如下:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

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

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

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

其中, 准确率用于衡量模型整体分类正确率; 精确率反映模型预测为呼吸暂停事件的样本中真实为正类的比例; 召回率用于衡量模型对实际呼吸暂停事件的检出能力; F1 值综合考虑精确率与召回率, 能够更全面地评价模型在类别不平衡条件下对正类事件的识别性能。

4.3. 对比实验

Table 5. Performance comparison between the proposed model and other models
表 5. 本文模型与其他模型对比

方法	时间粒度	Accuracy (%)	Precision (%)	Recall (%)
SomnNET [6]	1 s	70.85	75.62	65.31
SomnNET [6]	60 s	72.43	78.91	68.74
Bi-LSTM-CNN [7]	60 s	78.12	80.45	74.63
1D-Unet	1 s	72.09	75.99	58.41
Transformer [8]	1 s	77.32	78.63	73.46
本文模型	1 s	79.96	76.77	78.00

为验证本文所提出模型的有效性, 本文选取了 SomnNET [6]、Bi-LSTM-CNN [7]以及 1D-Unet 和 Transformer [8]等代表性方法作为对比模型, 并在相同数据集和统一实验设置下进行了性能比较。各方法分别采用不同时间粒度进行呼吸暂停检测, 其中 SomnNET 和 1D-Unet 主要在单一时间尺度下进行建模, 而本文方法采用多粒度联合建模策略, 在不同时间分辨率下同时学习呼吸暂停事件的局部细节与整体演化特征。

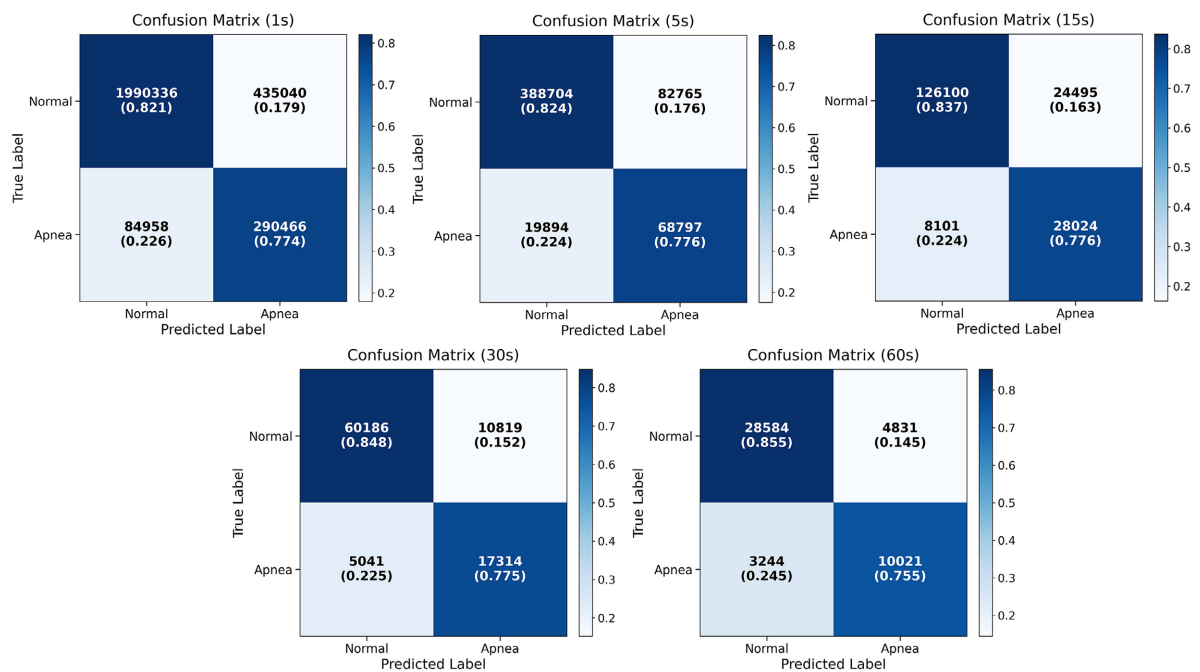


Figure 3. Confusion matrices at different temporal granularities
图 3. 不同粒度下的混淆矩阵

表 5 给出了不同模型的对比结果。从实验结果可以看出, 本文模型在综合性能上优于其他对比方法。其中, 本文模型的准确率达到 79.96%, 为所有方法中最高; 召回率达到 78.00%, 同样优于其余对比方法, 说明本文方法能够更有效地识别呼吸暂停事件, 降低漏检率。虽然在精确率指标上, Bi-LSTM-CNN [7]取得了最高的 80.45%, 略高于本文模型的 76.77%, 但其召回率仅为 74.63%, 整体均衡性不及本文方法。相比之下, 本文模型在准确率与召回率之间取得了更优的折中, 体现出更稳定的检测性能。

与上述方法相比, 本文模型通过引入多粒度建模机制, 在细粒度和粗粒度两个层面同时提取特征,

使模型既能够捕捉短时呼吸暂停事件的局部变化, 又能够建模长时间范围内的整体呼吸模式, 因此在准确率和召回率指标上均表现出明显优势。实验结果表明, 多粒度联合建模对于提升血氧信号呼吸暂停检测性能具有重要作用, 也验证了本文所提方法的有效性。

为了进一步分析模型在不同时间粒度下的分类表现, 本文绘制了 1 s、5 s、15 s、30 s 和 60 s 五种时间粒度下对应任务头的混淆矩阵, 如图 3 所示。可以看出, 在不同粒度设置下, 模型对正常片段和呼吸暂停片段均表现出较好的区分能力。随着时间粒度逐渐增大, 正常类别的识别准确性有所提升。这说明较粗粒度的时间建模能够更好地提取呼吸信号在较长时间范围内的整体变化趋势, 从而减少正常片段的误判。

4.4. 稳定性实验

为验证模型结果的稳定性与泛化能力, 本文采用五折交叉验证策略进行实验。具体而言, 将全部受试者样本划分为 5 个互不重叠的子集, 每次选取其中 1 个子集作为测试集, 其余 4 个子集用于模型训练与验证; 重复进行 5 次实验, 使每个子集均恰好作为 1 次测试集。最终以 5 次实验结果的平均值和标准差作为模型的总体性能评价指标。实验中采用 Accuracy、Precision、Recall 和 F1-Score 作为主要评价指标。

表 6 给出了本文模型在五折交叉验证下的实验结果。可以看出, 模型在不同折次上的性能波动较小, 说明本文方法具有较好的稳定性和泛化能力。具体而言, 5 折实验的准确率分别为 80.33%、80.70%、79.89%、80.08% 和 78.82%, 平均值达到 79.96%, 标准差仅为 0.63。与此同时, Precision、Recall 和 F1-Score 的平均值分别为 76.77%、78.03% 和 77.39%, 对应标准差分别为 1.28、0.74 和 0.51。整体来看, 各折结果分布较为集中, 表明所提出模型在不同数据划分下均能够保持稳定的检测性能。

Table 6. Results of five-fold cross-validation

表 6. 五折交叉验证结果

折次	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Fold 1	80.33	77.62	78.48	78.05
Fold 2	80.70	78.33	76.95	77.64
Fold 3	79.89	76.85	77.34	77.09
Fold 4	80.08	76.53	78.67	77.59
Fold 5	78.82	74.53	78.72	76.57
Mean \pm Std	79.96 \pm 0.63	76.77 \pm 1.28	78.03 \pm 0.74	77.39 \pm 0.51

4.5. 可视化实验

为进一步分析所提多时间粒度深监督模型在不同时间粒度下的决策依据, 本文采用 Grad-CAM 对典型 SpO₂ 样本进行可视化分析, 观察各粒度监督分支在输入序列上的关注区域。图 4 给出了 1 s、5 s、15 s、30 s 和 60 s 五个时间粒度下的 Grad-CAM 响应结果, 其中黑色曲线表示原始 SpO₂ 信号, 背景颜色表示模型对相应时间区域的响应强度。

从图 4 可以看出, 不同时间粒度分支在同一样本上的关注模式存在明显差异。细粒度分支(1 s 和 5 s)对 SpO₂ 信号中的局部下降边沿、短时波动及恢复转折表现出更强的响应, 说明其更关注呼吸事件发生过程中的瞬时变化和边界信息。相比之下, 较粗粒度分支(30 s 和 60 s)的高响应区域分布更为平滑, 主要覆盖持续时间较长的低氧变化区间, 表明其更倾向于利用较长时间范围内的上下文信息对异常片段进行判断。15 s 分支则在二者之间起到过渡作用, 既能反映局部异常变化, 又在一定程度上保留了持续模式信息。

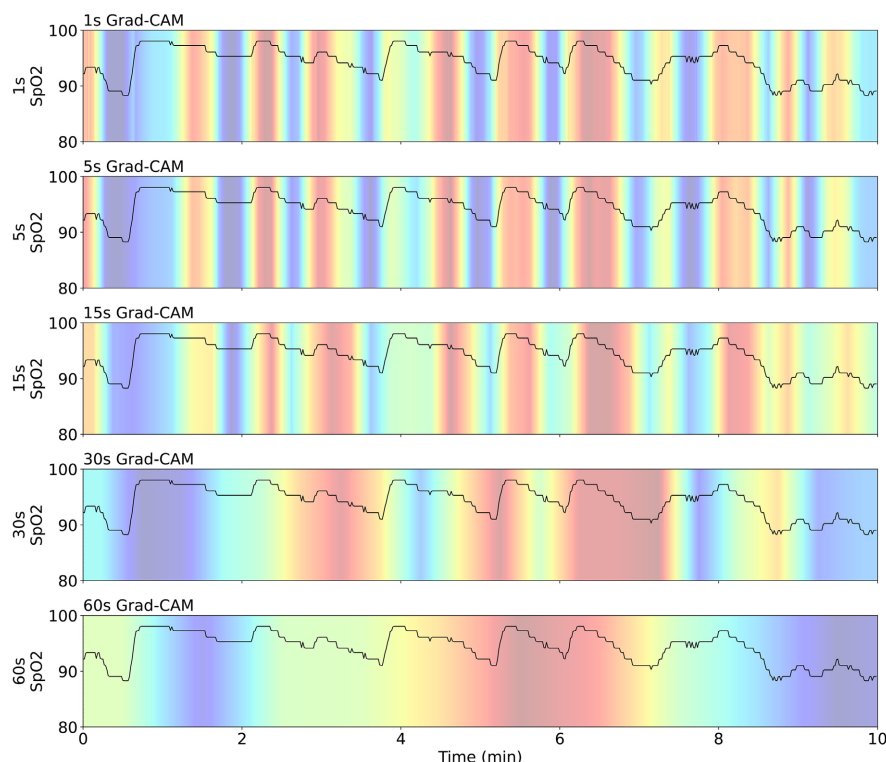


Figure 4. Visualization results of Grad-CAM at different time granularities
图 4. 不同时间粒度下的 Grad-CAM 可视化结果

4.6. 消融实验

为验证不同时间粒度监督分支对模型性能的影响, 本文进一步开展逐步累积的粒度消融实验。消融实验中以 1 s 粒度输出作为主任务结果, 并在此基础上逐步引入 5 s、15 s、30 s 和 60 s 辅助监督分支, 以分析不同时间粒度信息对睡眠呼吸暂停检测性能的贡献。除监督分支设置外, 其余实验条件均与主实验保持一致, 包括数据划分方式、输入设置、训练参数及评价指标, 从而保证不同模型之间比较的公平性。

本文共设计 5 组消融模型, 具体如下:

- 1) 基线: 仅保留 1 s 主任务分支, 用于构建单一时间粒度检测基线。
- 2) 模型 1: 在基线基础上增加 5 s 辅助监督分支, 以增强模型对短时局部上下文信息的建模能力。
- 3) 模型 2: 在 Case 2 基础上进一步增加 15 s 辅助监督分支, 以补充中短时范围内的事件结构信息。
- 4) 模型 3: 在 Case 3 基础上继续增加 30 s 辅助监督分支, 以增强模型对较长时间持续模式的刻画能力。
- 5) 完整模型: 在 Case 4 基础上进一步增加 60 s 辅助监督分支, 构成完整的多时间粒度深监督模型, 用于同时建模局部短时变化与长时间持续异常模式。

从表 7 可以看出, 随着监督粒度由细到粗逐步增加, 模型性能整体呈现逐步提升趋势, 说明多时间粒度监督能够有效增强模型对呼吸暂停事件层次化时序特征的建模能力。具体而言, 基线仅采用 1 s 主任务监督, 能够获得较为稳定的基线性能, 但对较长时间范围内事件持续模式的刻画能力有限。引入 5 s 辅助监督后, 模型 1 的性能较基线有所提升, 表明短时辅助粒度有助于增强模型对局部动态变化的识别能力。在此基础上, 模型 2 和模型 3 分别进一步引入 15 s 和 30 s 监督分支, 模型性能持续提高, 说明多层次时间粒度信息能够在局部细节与中长时上下文之间形成有效互补。最终, 完整模型在引入 60 s 粒度监督后取得最优结果, 表明较粗粒度分支能够进一步增强模型对长时间持续异常模式和事件聚集特征的代表能力。

Table 7. Ablation study results
表 7. 消融实验结果

模型	1 s	5 s	15 s	30 s	60 s	Accuracy (%)	F1-Score (%)
基线	√	×	×	×	×	72.09	66.12
模型 1	√	√	×	×	×	76.64	70.36
模型 2	√	√	√	×	×	77.76	75.11
模型 3	√	√	√	√	×	78.11	75.50
完整模型	√	√	√	√	√	80.59	80.18

5. 结论

本文围绕 SpO₂ 信号在睡眠呼吸暂停检测中的应用需求, 提出了一种基于多时间粒度深监督的单模态检测方法。该方法以一维 U-Net 为主体结构, 引入多时间粒度标签构建、多粒度分类头以及深监督联合优化策略, 在统一框架下同时建模逐秒级局部低氧变化与较长时间范围内的事件持续模式。实验结果表明, 所提方法在五折交叉验证中表现稳定, 说明多时间粒度建模有助于提升 SpO₂ 信号对呼吸暂停事件的层次化表征能力。该研究验证了血氧信号在低侵入式睡眠呼吸暂停检测任务中的重要价值, 并为后续多模态融合建模及智能睡眠监测系统呼吸异常分析提供了方法基础。

参考文献

- [1] Terrill, P.I. (2020) A Review of Approaches for Analysing Obstructive Sleep Apnoea-Related Patterns in Pulse Oximetry Data. *Respirology*, **25**, 475-485. <https://doi.org/10.1111/resp.13635>
- [2] Hoang, N.H. and Liang, Z. (2025) Detection and Severity Classification of Sleep Apnea Using Continuous Wearable SpO₂ Signals: A Multi-Scale Feature Approach. *Sensors*, **25**, Article 1698. <https://doi.org/10.3390/s25061698>
- [3] Rolon, R.E., Gareis, I.E., Larrateguy, L.D., Di Persia, L.E., Spies, R.D. and Rufiner, H.L. (2020) Automatic Scoring of Apnea and Hypopnea Events Using Blood Oxygen Saturation Signals. *Biomedical Signal Processing and Control*, **62**, Article ID: 102062. <https://doi.org/10.1016/j.bspc.2020.102062>
- [4] Liu, R., Li, C., Xu, H., Wu, K., Li, X., Liu, Y., et al. (2022) Fusion of Whole Night Features and Desaturation Segments Combined with Feature Extraction for Event-Level Screening of Sleep-Disordered Breathing. *Nature and Science of Sleep*, **14**, 927-940. <https://doi.org/10.2147/nss.s355369>
- [5] Sharma, M., Kumbhani, D., Yadav, A. and Acharya, U.R. (2022) Automated Sleep Apnea Detection Using Optimal Duration-Frequency Concentrated Wavelet-Based Features of Pulse Oximetry Signals. *Applied Intelligence*, **52**, 1325-1337. <https://doi.org/10.1007/s10489-021-02422-2>
- [6] John, A., Nundy, K.K., Cardiff, B. and John, D. (2021) SomnNET: An SpO₂ Based Deep Learning Network for Sleep Apnea Detection in Smartwatches. 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico, 1-5 November 2021, 1961-1964. <https://doi.org/10.1109/embc46164.2021.9631037>
- [7] Deviaene, M., Testelmans, D., Buyse, B., Borzee, P., Van Huffel, S. and Varon, C. (2019) Automatic Screening of Sleep Apnea Patients Based on the SpO₂ Signal. *IEEE Journal of Biomedical and Health Informatics*, **23**, 607-617. <https://doi.org/10.1109/jbhi.2018.2817368>
- [8] Almarshad, M.A., Al-Ahmadi, S., Islam, M.S., BaHammam, A.S. and Soudani, A. (2023) Adoption of Transformer Neural Network to Improve the Diagnostic Performance of Oximetry for Obstructive Sleep Apnea. *Sensors*, **23**, Article 7924. <https://doi.org/10.3390/s23187924>
- [9] Hou, Y., Wang, B., Zhang, C., Wang, Q., Li, J., Meng, P., et al. (2025) OSASformer: A Transformer-Based Model for OSAS Screening via Multi-Source Representation Fusion. *Knowledge-Based Systems*, **316**, Article ID: 113365. <https://doi.org/10.1016/j.knosys.2025.113365>
- [10] Hoang, N.H. and Liang, Z. (2025) AI-Driven Sleep Apnea Screening with Overnight Blood Oxygen Saturation: Current Practices and Future Directions. *Frontiers in Digital Health*, **7**, Article 1510166. <https://doi.org/10.3389/fgdh.2025.1510166>
- [11] Quan, S.F., Howard, B.V., Iber, C., et al. (1997) The Sleep Heart Health Study: Design, Rationale, and Methods. *Sleep*, **20**, 1077-1085.