

使用来自 EOG 和 PSM 的多模态嵌入融合进行睡眠阶段分类

Olivier Papillon
Systems and Computer Engineering
Carleton University
olypapillon@cmail.carleton.ca

Rafik Goubran, Life Fellow IEEE
Systems and Computer Engineering
Carleton University
Bruyère Health Research Institute
goubran@sce.carleton.ca

James Green, Senior Member IEEE
Systems and Computer Engineering
Carleton University
jrgreen@sce.carleton.ca

Julien Larivière-Chartier
Bruyère Health Research Institute
Systems and Computer Engineering
Carleton University
JLariviereChartier@bruyere.org

Caitlin Higginson
Sleep Research Unit
University of Ottawa Institute for Mental
Health Research at the Royal Ottawa Hospital
Caitlin.Higginson@theroyal.ca

Frank Knoefel, MD
Systems and Computer Engineering
Carleton University
Bruyère Health Research Institute
fknoefel@bruyere.org

Rébecca Robillard
School of Psychology,
University of Ottawa
University of Ottawa Institute for Mental
Health Research at the Royal Ottawa Hospital
Rebecca.Robillard@uottawa.ca

Abstract—准确的睡眠阶段分类对于诊断睡眠障碍特别是老年群体中的睡眠障碍至关重要。虽然传统的多导睡眠图 (PSG) 依赖脑电图 (EEG) 作为金标准, 但其复杂性和对专业设备的需求使得居家睡眠监测具有挑战性。为了解决这一限制, 我们研究了使用眼动电图 (EOG) 和压力敏感垫 (PSM) 作为五阶段睡眠觉醒分类的较不显眼的替代方案。本研究介绍了一种新的方法, 利用多模态嵌入深度学习模型 ImageBind, 将 PSM 数据与双通道 EOG 信号集成用于睡眠阶段分类。我们的方法是首次报道使用 ImageBind 融合 PSM 和 EOG 数据进行睡眠阶段分类的方法。我们的结果表明, 微调 ImageBind 显著提高了分类准确性, 优于基于单通道 EOG (DeepSleepNet)、仅使用 PSM 数据 (ViViT) 以及其他多模态深度学习 (MBT) 的现有模型。值得注意的是, 该模型在无需微调的情况下也取得了良好的性能, 强调了其在有限标签数据的特定任务中的适应性, 使其在医疗应用中尤为有利。我们使用来自睡眠诊所的患者 85 个夜晚的记录评估了我们的方法。我们的研究结果表明, 预训练的多模态嵌入模型, 即使是那些最初为非医疗领域开发的模型, 也可以有效地适应睡眠分期, 其准确度接近需要复合 EEG 数据的系统。

Index Terms—ImageBind, sleep stages, multimodal, embeddings, EOG, PSM, classification, deep learning, pre-training, fine-tuning

I. 介绍

睡眠障碍影响着大量的加拿大人, 并且对身体和心理健康的多个方面产生重要影响。然而, 获取睡眠测试一直存在长期问题, 成为高效睡眠医疗的障碍。远程健康监控可

以通过减少对病床的需求、解决长候补名单和高住院成本等问题, 缓解加拿大医疗系统的压力。[1]。

多导睡眠图 (PSG) 是对睡眠的研究, 随着年龄增长, 睡眠相关障碍变得更加普遍, 因此变得越来越重要。睡眠阶段分类是识别个体所处不同睡眠阶段的过程。研究表明, 睡眠障碍在特定睡眠阶段加剧; 例如, 睡眠呼吸暂停在浅睡 1 阶段 (NREM1)、浅睡 2 阶段 (NREM2) 和快速眼动阶段 (REM) 更为常见。[2]。准确的睡眠阶段分类使研究人员能够为下游任务 (如自动睡眠呼吸暂停事件检测) 选择最合适的估计方法。目前, 自动睡眠阶段分类的最新技术使用来自电极位置在 FpZ (前额中线位置) 和 Cz (中央中线位置) 之间的差分信号的单通道脑电图 (EEG) 数据。[3]。然而, 尽管性能很高, 使用基于 EEG 的传感器仍存在一些实际挑战, 包括需要经过培训的临床医生、为了降低皮肤电极电阻进行的大量皮肤准备, 以及提供受控环境的高级实验室设施。由于 EEG 传感器的复杂性, 它们对远程健康监控构成挑战。因此, 我们专注于侵入性较小的传感器模式, 如压力敏感垫 (PSMs) 和眼电图 (EOG)。PSMs 测量患者与床垫之间在空间和时间上的接触压力。PSMs 已被证明在多种健康监测应用中是有效的, 包括睡眠呼吸暂停检测 [4], 生命体征估计 [5], [6], 患者运动检测 [7]–[9], 误报检测 [10] 和睡眠阶段分类 [11], [12]。EOG 信号可以通过放置在眼外眦和内眦附近皮肤上的电极来测量, 表示在眼球运动过程中角膜和视网膜之间的电位变化。眼球运动的准确记录是有价值的, 因为滚动眼球

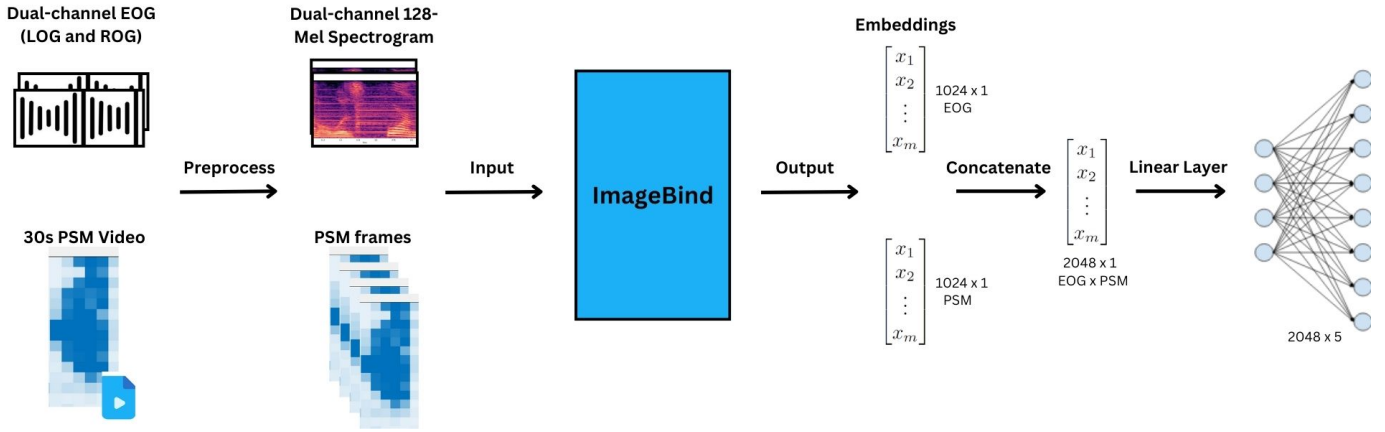


Fig. 1: 眼动图 (EOG) 和姿势监测 (PSM) 多模态嵌入融合过程的模型概述

运动表示从清醒到浅睡眠的过渡，而快速眼球运动则表征了快速眼动期 (REM) 睡眠 [13]。此外，它已被应用于疲劳追踪 [14]，使用眼动追踪面罩进行睡眠监测 [15]，以及睡眠阶段分类 [16]–[18]。研究人员已经提议使用不那么显眼的传感器模式和更易于在家庭环境中部署的传感器来对患者进行有效的睡眠监测。[19] 在这些模式中，EOG 和 PSM 作为不显眼且更实用的选项，在基于家庭的睡眠监测系统中脱颖而出。

在本研究中，我们专注于使用 PSM 和 EOG 数据自动识别多导睡眠图记录的睡眠阶段。为实现这一目标，我们使用机器学习 (ML) 融合并分析 PSM 和 EOG 数据，提取有意义的模式和有价值的见解。最近的研究已经通过深度学习架构 [16], [18], [20] 利用 EOG 确定睡眠阶段。虽然这些方法已被证明有效，但最近整合多模态的深度学习架构在机器学习任务中显示出改进的分类性能。[21] 我们假设 PSM 数据 (捕捉较大身体运动) 和 EOG (捕捉眼球运动) 的融合比单独的模态更为有效。由于 PSM 数据表示时间上的图像 (即视频)，而 EOG 信号是生物电信号 (即类似音频信号)，我们的数据可以表示为视频和音频信号的结合，因此可以使用旨在融合音视频数据的深度学习模型 [22], [23]。

虽然深度学习方法非常强大，但由于需要大量标记数据，训练这些模型仍然是一个重大挑战 [24]。为了克服当前标记训练数据的短缺，我们利用迁移学习来应用最新的预训练深度学习模型，并针对我们的特定任务进行微调。在本文中，我们评估了基于 EOG 的睡眠阶段分类的最新技术，使用来自实际患者的数据；我们证明了 EOG 数据的 ImageBind 嵌入在睡眠阶段分类模型中优于现有技术；最终，我们展示了使用 ImageBind 多模态嵌入融合 EOG 和 PSM 数据相对于这些模型可以进一步提高睡眠阶段分类的准确性。这项研究通过研究 EOG 和 PSM 数据在检测睡眠阶段的有效性，为计算机视觉、机器学习和医学测量及应用领域做出了贡献。PSM 和 EOG 的结合需要最少的安装并尊重用户隐私。这些特性使得它们非常适合用于为远程患者监控创建非侵入性的睡眠跟踪解决方案。

在 [8] 中，Foubert 等人使用经典的机器学习 (ML) 方法进行睡眠姿势分类。研究人员从 PSM 数据中提取特征 (例如，传感器值的总和和压力中心) 来训练 SVM 和 K-NN 分类器，以识别六种不同的睡眠姿势。在 [12] 中，Green

等人还利用 PSM 数据通过使用 TCN 网络来识别睡眠与清醒状态。

在

A. 多模态睡眠阶段分类

Lin 等人开发了一种多模态睡眠阶段分类方法，使用来自两个数据集 (SleepEDF-ST 和 SleepEDF-SC) 中 100 名参与者的 EOG 和 EEG 数据 [25]。该研究采用了一种混合深度学习架构，结合了多模态局部特征提取、整体相关编码和跨域交互模块，以分类五个睡眠-觉醒阶段 (Wake, N1, N2, N3 和 REM)。

B. 使用 EOG 数据的睡眠阶段分类

在 [18] 中，Maiti 等人提出了一种新的 SE-Resnet-Transformer 模型，用于使用单通道 EOG 数据进行睡眠阶段分类。该研究利用 EOG 模态进行自动分类五个睡眠觉醒阶段 (醒觉、N1、N2、N3 和 REM)。在 [16] 中，Jiahao 等人介绍了 EOGNet，这是一种设计用于使用 EOG 数据进行睡眠阶段分类的深度学习模型。该模型结合了两级卷积神经网络 (CNN) 进行特征提取，以及双向门控循环单元 (Bi-GRU) 捕捉睡眠时期的序列信息。采用两步训练协议：首先在平衡数据集上对 CNN 进行预训练，然后用序列输入对整个模型进行微调。

在音视频深度学习方面，Nagrani 等人在 [23] 中引入了一种基于 transformer 的模型，称为多模态瓶颈 Transformer (MBT)，以改进视频分类中的音视频融合。传统模型依赖于后期融合，而 MBT 则使用一小组潜在单元形成注意力瓶颈，迫使模型在给定层内在分享信息之前，从每种模式中整理和压缩信息，从而在提高性能的同时降低计算成本。在 [26] 中，Arnab 等人为视频分类引入了基于 transformer 的模型，用自注意力取代卷积神经网络以捕捉长距离依赖性。通过利用在大型图像数据集上的预训练，他们在相对较小的视频数据集上训练了 ViViT。在诸如 Kinetics 400 [27] 的基准上进行评估，ViViT 取得了最先进的性能。

在 [22] 中，Girdhar 等人引入了一种模型，该模型学习了一个跨越六种不同模态的联合多模态嵌入空间：图像、文本、音频、深度、热成像和 IMU 数据。与之前需要显式多模态对数据来进行训练的方法不同，ImageBind 仅利

用图像配对数据,通过使用图像的绑定特性将所有模式对齐到一个共享表示中。在目前的研究中,我们微调了 ImageBind 以对齐 EOG 和 PSM 数据,并使用其进行睡眠阶段分类。

C. DeepSleepNet

年,Supratak 等人在 [20] 中提出了一种基于深度学习的模型,用于使用原始单通道 EEG 或 EOG 数据进行自动睡眠阶段分类。与以前依赖于人工特征工程的方法不同,DeepSleepNet 采用卷积神经网络 (CNN) 进行时间不变特征提取,并使用双向长短期记忆 (Bi-LSTM) 网络捕捉时间依赖性,包括睡眠阶段转换。它被广泛认为是睡眠阶段分类领域的最先进的模型。在本研究中,该方法作为基线,用于比较我们多模式睡眠分类模型的有效性。

II. 方法

A. 临床数据收集

在渥太华皇家医院的睡眠障碍诊所进行了一项过夜研究,安装了一套包括脑电图 (EEG)、肌电图 (EMG)、眼电图 (EOG) 和床基压力传感垫 (PSM) 的设备,这些设备记录整夜的接触压力数据,以识别睡眠研究期间的不同睡眠阶段。研究中,参与者在两个房间内使用 PSM 及标准多导睡眠监测 (PSG) 传感器 (包括 EEG、EOG 和 EMG) 进行监测。睡眠数据被连续记录并随后由注册的睡眠技师根据标准标准 [28] 对睡眠姿势 (四种姿势) 和睡眠阶段 (五个阶段) 信息进行标注。EOG 传感器记录患者的眼球运动,每只眼睛 (左眼和右眼) 各配置一个传感器。这些传感器输出的数值范围从 -3000 到 3000,然后被归一化到 0 至 1 的刻度。我们的数据集包括 85 个夜晚的患者数据记录。对于前 74 位参与者,数据以 250 Hz 采样;由于医院设备的升级,剩余 11 位参与者的采样率提高到 512 Hz。诊所中的每张床都装有由 Tactex Controls Inc. 制造的多个 PSM。这些垫由 3×8 光纤接近传感器网格组成,能以 10 Hz 的采样率检测压力。传感器的数值范围从 0 到 2046,被归一化到 0 至 1 的刻度。每个房间有三个 PSM 连接到用于上半身的数据采集箱 (DAQ),还有三个 PSM 用于中下半身连接到其他 DAQ,生成 18×8 的低分辨率图像输出。该数据集包括五个睡眠-觉醒阶段:醒、非快速眼动 1 期 (NREM1)、非快速眼动 2 期 (NREM2)、非快速眼动 3 期 (NREM3) 和快速眼动 (REM)。

我们的数据集包含 85 名受试者的过夜睡眠记录,平均每晚睡眠时长约为 6.5 小时。每个 30 秒的时段都有睡眠阶段标签;因此,PSM 和 EOG 数据也被分割成不重叠的 30 秒时段,假设在每个时段内睡眠阶段保持不变。这大约导致每位患者 780 个时段。结合 85 名研究参与者,我们的最终数据集包含 63,236 个时段。PSM 数据是以 10 Hz 采集的。EOG 信号的采样率为 250 Hz 或 512 Hz (因患者不同而异)。为了通过 ImageBind 进行编码,EOG 数据被存储为双通道 WAV 音频文件,左右声道分别对应文件中的相应通道。因此,数据集中的每条记录对应一个 30 秒时段,并包含一个 PSM 视频和一个双通道 EOG wav 文件。

为了进行模型评估,我们采用了五折交叉验证方案以确保稳健性。每折包含大约 17 名患者,因此每折的训练集有 68 名患者,测试集有 17 名患者。来自同一名患者的数据从未同时用于训练和测试,从而避免了过拟合

为了融合 EOG 和 PSM 模式,我们使用 ImageBind 嵌入对其进行对齐。ImageBind 模型使用共享的嵌入空间将多种模式——包括图像、音频、文本、热成像、深度和 IMU 信号——绑定到图像对表示。首先,每种模式的数据经过预处理,以成为 ImageBind 可接受的输入。对于 EOG, WAV 文件被上采样到 16 kHz,然后转换为 128 梅尔频谱图。对于 PSM,每个 30 秒的视频被转换为单个帧进行处理。然后,这些模式通过其指定的模式 (即,对于 EOG 是音频,对于 PSM 是视频) 输入到预训练的 ImageBind 模型中。我们评估了两种配置:一种是 ImageBind 的线性探针 (即权重保持不变),另一种是对 ImageBind 进行微调的配置。模型为每种模式输出一个大小为 1024×1 的向量 (或嵌入)。然后,我们将这两个向量连接起来,形成一个大小为 2048×1 的向量,并添加一个线性层以将融合的代表映射到五类分类。这在图 1 中有所说明。

我们采用了几个预训练模型,这些模型运行在单一或多个模式上。

如前所述,我们使用了两个 ImageBind 模型:一个是在两个输入上进行微调的模型,另一个是权重被冻结的 (线性探测)。此外,我们还采用了一个仅处理 EOG 输入的 ImageBind 模型。(图 ??)。在此配置中,输出向量 (1024×1) 被输入到维度为 1024×5 的线性层中。

ViViT 模型 (图 ??) 是一种在 Kinetics-400 [27] 数据集上预训练的最先进模型,该数据集包含 400 个 10 秒的 RGB 视频片段,涵盖 400 个类别,分辨率为 224×224 像素。ViViT 被设计用于处理视频片段,因此我们使用 PSM 视频对其进行了训练。然而,PSM 产生的数据是单通道、分辨率仅为 18×8 像素的归一化压力读数。为了使数据与模型兼容,我们修改了变换器的位置嵌入,以支持较小的视频输入大小,并将三通道的 RGB 输入权重加和为一个通道以适应单通道的 PSM 数据。

MBT 模型 (图 ??) 是一种最先进的多模式模型,它采用了两个主干模型,一个用于视频,一个用于音频。视频模型是一个在 ImageNet-21k [29] 上预训练的 ViT 模型,其中包含了 21,843 个类别的 1400 万张图像,RGB 图像的分辨率为 224×224 像素。音频模型同样是一个在 AudioSet [30] 上预训练的 ViT 模型,由人类标注的 10 秒声音片段组成了 632 个音频类别,它的输入是音频片段的 128 梅尔频谱图。为了使数据与 MBT 兼容,我们将 PSM 视频转换为 RGB 格式,并使用双线性插值方法上采样至 224×224 像素,同时将 LOG (左眼动电图) 的单一通道转换为 128 梅尔频谱图。

在基于 EOG 的睡眠阶段分类排行榜上,排名前三的方法是 EOGNet [16]、SeResNet [18] 和 DeepSleepNet [20]。虽然我们尝试联系前两种方法的作者,但未能重现其结果。然而,由于 DeepSleepNet 代表了基于 EOG 的睡眠阶段分类的最新技术,我们将其纳入了我们的性能比较。对于 DeepSleepNet,我们对模型进行了临床数据集的微调。因为 DeepSleepNet 需要 3,000 个样本和单通道的输入,因此我们对 LOG 信号进行了下采样。

B. 模型开发

所有模型均在使用 PyTorch 的 A100 GPU 上运行。对学习率、权重衰减和训练周期数进行了超参数搜索,以确定在五次折叠中产生最高平均五分类准确度的配置。所有模型均在我们的临床数据集上进行了微调。

TABLE I: 我们临床数据集上睡眠阶段分类结果的比较

Method	Modality	Accuracy	F1 (macro)
ViVit	PSM video	0.399	0.164
MBT	PSM video & Single-channel EOG	0.631	0.543
DeepSleepNet	Single-channel EOG	0.743	0.682
ImageBind	PSM video & Dual-channel EOG	0.745	0.683

我们比较了不同的机器学习架构在仅使用 PSM、单通道 EOG 以及两种模态的情况下进行睡眠阶段分类的表现。在五次交叉验证中，五分类的准确率和宏观平均 F1 得分被计算平均。虽然 K 折交叉验证确保相同的数据不会同时用于训练和测试模型，但在超参数优化过程中重复交叉验证有时会导致过拟合。这里开发的模型将在未来的数据收集过程中对新病人的数据进行进一步验证。

来自表格 I 的结果表明，微调 ImageBind 并使用 PSM 和双通道 EOG 的效果优于所有其他方法。值得注意的是，仅使用 PSM 视频的模型表现不如其他模态。然而，仅引入 EOG 传感器模态就显著改善了模型的性能。这种改进可以归因于睡眠阶段与 EOG 信号之间的强相关性，特别是在区分 REM 和 NREM 阶段时。此外，通过结合 PSM 与正确的 EOG 来包含多种模态，提供了额外的特征，进一步增强了睡眠阶段的分类。这些发现表明，多模态方法如 ImageBind 可以通过有效利用来自不同数据模态的互补信息，显著改善睡眠阶段的分类。

我们性能最佳的 ImageBind 模型使用 PSM 和双通道 EOG 的分类模型，其最优超参数如下：我们使用了一个学习率调度器，其初始学习率为 $1.4e-7$ ，权重衰减为 0.005，批大小为 12。我们使用 AdamW 优化器进行微调，共进行了 6 个周期。

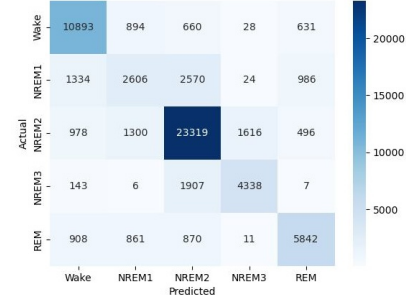
C. ImageBind 训练方法的比较

我们使用两种不同的方法训练了 ImageBind，如表 II 所示。首先，我们对 ImageBind 进行线性探测以评估其领域适应性，使我们能够评估它在不进行微调的情况下如何很好地对睡眠阶段进行分类。值得注意的是，ImageBind 即使在没有进行微调的情况下也表现良好，展示了其在有限标注数据情况下适应特定任务的能力，使其特别适用于医疗应用。这尤其令人印象深刻，因为模型处理 PSM 视频时，这些视频是低分辨率压力视频，并将 EOG 数据作为音频输入，这与其预期的输入模态不同。此外，我们观察到加入 PSM 视频数据的增量效益，准确率从 0.710 提升至 0.745，而 F1-score 从 0.636 提升至 0.683，突显了增加 PSM 数据在提升睡眠阶段分类性能方面的好处。

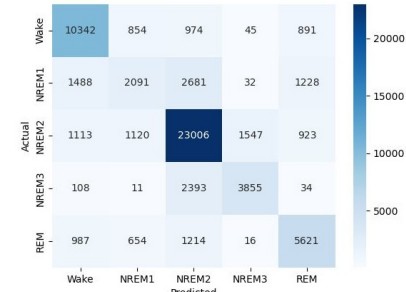
通过比较图 2a 和图 2b 的混淆矩阵，我们观察到两个模型在区分 REM、觉醒和 NREM2 阶段时表现良好。引入 PSM 数据进一步减少了分类错误，提高了整体模型的准确性。主要挑战仍然存在于区分 NREM2、NREM3 和 NREM1，这在仅依赖 EOG 和 PSM 数据时本质上是困难的。此外，大多数误分类发生在相邻的睡眠阶段之间（例如，NREM1 与觉醒和 NREM2，但不与 NREM3 或 REM）。这是可以预期的，因为睡眠阶段逐渐过渡，使得相邻阶段更难区分，而区分 REM 与觉醒相对更容易。我们的方法是首次报道使用 PSM 和 EOG 数据通过 ImageBind 进行睡眠阶段分类的方法。我们证明了即便在预训练任务（RGB 视频分类和声音分类）与目标领域（单通道低分辨率 PSM

TABLE II: 使用不同训练方法的图像绑定在我们的临床数据集上的睡眠阶段分类结果比较

Model	Method	Modality	Accuracy	F1 (macro)
ImageBind	Fine-tuned	Dual-channel EOG	0.710	0.636
ImageBind	Linear Probing	PSM video & Dual-channel EOG	0.690	0.614



(a) ImageBind PSM 和 EOG 的混淆矩阵



(b) 仅使用 ImageBind EOG 的混淆矩阵

Fig. 2: 使用两种输入模态的 ImageBind 的混淆矩阵: PSM 与 EOG 的融合，与仅 EOG

数据和多通道 EOG) 之间存在领域转移的情况下，使用不同模态或领域的预训练模型也能提高睡眠阶段分类的效果。这进一步说明了在医学应用中使用最先进的预训练模型的益处，特别是在可用数据有限的情况下。

III. 结论

我们的研究表明，使用多模态深度学习模型可以通过微调从实际患者在临床过夜睡眠研究中收集的数据来增强从 PSM 和多通道 EOG 数据进行睡眠阶段分类。我们显示了结合 PSM 可以改善有睡眠问题的个人的睡眠阶段分类。此外，该模型表现出强大的领域适应性，能够有效地从有限的标签数据中学习，这对医学应用尤其有利，因为标注数据集通常稀缺。这些结果表明，经过微调的 ImageBind 模型在家庭环境中进行实际睡眠阶段估计具有重大潜力。

在未来的工作中，我们将探讨位置估计和睡眠阶段分类的集成应用，例如睡眠呼吸暂停检测或生命体征估计，并在高分辨率 PSM 数据上验证结果。

我们感谢加拿大自然科学与工程研究委员会 (NSERC) 的支持。

F. Knoefel 感谢 Bruyère 健康研究院在居家养老技术研究主席方面的支持。

本研究遵循 IRB 批准的协议 # 112123“睡眠和健康改善；跟踪纵向变化 (SHIFT)”

REFERENCES

- [1] M. Cohen-McFarlane, S. Bennett, B. Wallace, R. Goubran, and F. Knoefel, “Bed-Based Health Monitoring Using Pressure Sensitive Technology: A Review,” *IEEE Instrumentation & Measurement Magazine*, vol. 24, no. 2, pp. 13–23, 2021.
- [2] C. Clinic, “Sleep Basics,” 2024. <https://my.clevelandclinic.org/12148-sleep-basics>.
- [3] E. Eldele, Z. Chen, C. Liu, M. Wu, C.-K. Kwok, X. Li, and C. Guan, “An Attention-Based Deep Learning Approach for Sleep Stage Classification With Single-Channel EEG,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 809–818, 2021.
- [4] H. Azimi, P. Xi, M. Bouchard, R. Goubran, and F. Knoefel, “Machine Learning-Based Automatic Detection of Central Sleep Apnea Events From a Pressure Sensitive Mat,” *IEEE Access*, vol. 8, pp. 173428–173439, 2020.
- [5] A. Bekele, S. Nizami, Y. S. Dosso, C. Aubertin, K. Greenwood, J. Harrold, and J. R. Green, “Real-time Neonatal Respiratory Rate Estimation using a Pressure-Sensitive Mat,” in *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 1–5, 2018.
- [6] H. Azimi, S. Soleimani Gilakjani, M. Bouchard, S. Bennett, R. A. Goubran, and F. Knoefel, “Breathing signal combining for respiration rate estimation in smart beds,” in *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 303–307, 2017.
- [7] N. Foubert, *Posture recognition and postural transition detection using bed-based pressure sensor arrays*. Master of Applied Science, Carleton University, Ottawa, Ontario, 2010.
- [8] N. Foubert, A. M. McKee, R. A. Goubran, and F. Knoefel, “Lying and sitting posture recognition and transition detection using a pressure sensor array,” in *2012 IEEE International Symposium on Medical Measurements and Applications Proceedings*, pp. 1–6, 2012.
- [9] S. Bennett, Z. Ren, R. Goubran, K. Rockwood, and F. Knoefel, “In-Bed Mobility Monitoring Using Pressure Sensors,” *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 8, pp. 2110–2120, 2015.
- [10] D. G. Kyrollos, K. Greenwood, J. Harrold, and J. R. Green, “Detection of False Alarms in the NICU Using Pressure Sensitive Mat,” in *2021 IEEE Sensors Applications Symposium (SAS)*, pp. 1–5, 2021.
- [11] L. Samy, M.-C. Huang, J. J. Liu, W. Xu, and M. Sarrafzadeh, “Unobtrusive Sleep Stage Identification Using a Pressure-Sensitive Bed Sheet,” *IEEE Sensors Journal*, vol. 14, no. 7, pp. 2092–2101, 2014.
- [12] C. Green, M. Bouchard, R. Goubran, R. Robillard, C. Higginson, E. Lee, and F. Knoefel, “Sleep-Wake and Body Position Classification with Deep Learning using Pressure Sensor Mat Measurements,” in *2023 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 1–6, 2023.
- [13] L. Smolley, “Adult polysomnography,” in *Encyclopedia of Sleep and Circadian Rhythms (Second Edition)* (C. A. Kushida, ed.), pp. 474–477, Oxford: Academic Press, second edition ed., 2023.
- [14] J. Xu, J. Min, and J. Hu, “Real-time eye tracking for the assessment of driver fatigue,” *Healthcare Technology Letters*, vol. 5, pp. 54–58, Jan. 2018.
- [15] S.-F. Liang, C.-E. Kuo, Y.-C. Lee, W.-C. Lin, Y.-C. Liu, P.-Y. Chen, F.-Y. Cherng, and F. Shaw, “Development of an EOG-Based Automatic Sleep-Monitoring Eye Mask,” *IEEE Transactions on Instrumentation and Measurement*, vol. 64, pp. 1–11, 2015.
- [16] J. Fan, C. Sun, M. Long, C. Chen, and W. Chen, “EOGNET: A Novel Deep Learning Model for Sleep Stage Classification Based on Single-Channel EOG Signal,” *Frontiers in Neuroscience*, vol. 15, 2021.
- [17] M. M. Rahman, M. I. H. Bhuiyan, and A. R. Hassan, “Sleep stage classification using Single-Channel EOG,” *Computers in Biology and Medicine*, vol. 102, pp. 211–220, 2018.
- [18] S. Maiti, S. K. Sharma, and R. S. Bapi, “Enhancing Healthcare with EOG: A Novel Approach to Sleep Stage Classification,” in *2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2305–2309, 2024.
- [19] Z. Huang, J. Li, and Z. He, “Full-coverage unobtrusive health monitoring of elders at homes,” *Internet of Things*, vol. 26, p. 101182, 2024.
- [20] A. Supratak, H. Dong, C. Wu, and Y. Guo, “DeepSleepNet: A Model for Automatic Sleep Stage Scoring Based on Raw Single-Channel EEG,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, p. 1998–2008, Nov. 2017.
- [21] Y. Li, M. E. H. Daho, P.-H. Conze, R. Zeghlache, H. L. Boité, R. Tadayoni, B. Cochener, M. Lamard, and G. Quellec, “A review of deep learning-based information fusion techniques for multimodal medical image classification,” 2024.
- [22] R. Girdhar, A. El-Nouby, Z. Liu, M. Singh, K. V. Alwala, A. Joulin, and I. Misra, “Imagebind: One Embedding Space To Bind Them All,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15180–15190, 2023.
- [23] A. Nagrani, S. Yang, A. Arnab, A. Jansen, C. Schmid, and C. Sun, “Attention Bottlenecks for Multimodal Fusion,” in *Advances in Neural Information Processing Systems* (M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, eds.), vol. 34, pp. 14200–14213, Curran Associates, Inc., 2021.
- [24] Y. Zhang, C. Li, L. Zhaoxia, and M. Li, “Semi-Supervised Disease Classification Based on Limited Medical Image Data,” *IEEE Journal of Biomedical and Health Informatics*, vol. PP, pp. 1–12, 01 2024.
- [25] B. Kemp, A. Zwinderman, B. Tuk, H. Kamphuisen, and J. Obery, “Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG,” *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 9, pp. 1185–1194, 2000.
- [26] A. Arnab, M. Dehghani, G. Heigold, C. Sun, M. Lučić, and C. Schmid, “ViViT: A Video Vision Transformer,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6836–6846, October 2021.
- [27] W. Kay, J. Carreira, K. Simonyan, B. Zhang, C. Hillier, S. Vijayanarasimhan, F. Viola, T. Green, T. Back, P. Natsev, et al., “The Kinetics Human Action Video Dataset,” *arXiv preprint arXiv:1705.06950*, 2017.
- [28] R. Berry, R. Brooks, S. Harding, R. Lloyd, S. Quan, M. Troester, and B. Vaughn, “Aasm scoring manual updates for 2017 (version 2.4),” *Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine*, vol. 13, 04 2017.
- [29] T. Ridnik, E. Ben-Baruch, A. Noy, and L. Zelnik-Manor, “Imagenet-21k pretraining for the masses,”
- [30] J. F. Gemmeke, D. P. W. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. C. Moore, M. Plakal, and M. Ritter, “Audio Set: An ontology and human-labeled dataset for audio events,” in *Proc. IEEE ICASSP 2017*, (New Orleans, LA), 2017.
- [31] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, “PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals,” *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000 (June 13).