



计算机应用
Journal of Computer Applications
ISSN 1001-9081, CN 51-1307/TP

《计算机应用》网络首发论文

题目： 基于通用手环的睡眠呼吸暂停检测
作者： 黄锦阳, 崔丰麒, 马长秀, 樊文东, 李萌, 李经宇, 孙晓, 黄林生, 刘志
收稿日期： 2024-09-02
网络首发日期： 2024-12-03
引用格式： 黄锦阳, 崔丰麒, 马长秀, 樊文东, 李萌, 李经宇, 孙晓, 黄林生, 刘志. 基于通用手环的睡眠呼吸暂停检测[J/OL]. 计算机应用.
<https://link.cnki.net/urlid/51.1307.TP.20241202.1503.009>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

基于通用手环的睡眠呼吸暂停检测

黄锦阳^{1,2}, 崔丰麒^{2,3,4}, 马长秀⁵, 樊文东¹, 李萌^{1*}, 李经宇⁴, 孙晓^{1,2,4}, 黄林生⁶, 刘志⁷

(1. 合肥工业大学 计算机与信息学院, 合肥 230009; 2. 情感计算与先进智能机器安徽省重点实验室, 合肥 230009;

3. 中国科学技术大学 先进技术研究院, 合肥 230031;

4. 合肥综合性国家科学中心 人工智能研究院(安徽省人工智能实验室)合肥 230092;

5. 安徽医科大学第二附属医院 呼吸与危重症医学科, 合肥 230601; 6. 安徽大学 互联网学院, 合肥 230039;

7. 电气通信大学, 日本 东京 1828585)

(* 通信作者电子邮箱 mengli@hfut.edu.cn)

摘要: 睡眠呼吸暂停严重影响生活质量和健康。多导睡眠图(PSG)是诊断的“金标准”, 但成本高且不便长期监测。基于此, 一种基于通用运动手环的新方法被提出以便捷化检测睡眠呼吸暂停。通过分析手环采集的心率、血氧饱和度和睡眠状态数据, 采用自适应生理数据重构方法和数据插值方法, 实现噪声滤除。在特征工程中, 融合连续生理变量和类别变量, 以深度提取睡眠状态特征。分类模块采用轻量级门控循环单元模型, 简化训练过程, 降低过拟合风险。实验在23人数据集上获得93.68%准确率和93.97%召回率。相关性分析发现血氧饱和度、身体质量指数和年龄被确认为是判断睡眠呼吸暂停的关键特征。与多导仪相比, 所提方法更适用于家庭环境下的长期监测。

关键词: 睡眠呼吸暂停检测; 通用手环; 多模态数据处理; 长时健康监测; 多因素影响指标分析

中图分类号: TP391 文献标志码: A

Sleep apnea detection based on universal wristband

HUANG Jinyang^{1,2}, CUI Fengqi^{2,3,4}, MA Changxiu⁵, FAN Wendong¹, LI Meng^{1*},

LI Jingyu⁴, SUN Xiao^{1,2,4}, HUANG Linsheng⁶, LIU Zhi⁷

(1. School of Computer and Information, Hefei University of Technology, Hefei Anhui 230009, China;

2. Anhui Province Key Laboratory of Affective Computing & Advanced Intelligent Machine, Hefei Anhui 230009, China;

3. Institute of Advanced Technology, University of Science and Technology of China, Hefei Anhui 230031, China;

4. Artificial Intelligence Research Institute, Hefei Comprehensive National Science Center (Anhui Artificial Intelligence Laboratory), Hefei Anhui 230092, China;

5. Department of Respiratory and Critical Care Medicine, The Second Affiliated Hospital of Anhui Medical University, Hefei Anhui 230601, China;

6. School of Internet, Anhui University, Hefei Anhui 230039, China; 7. The University of Electro-Communications, Tokyo, 1828585, Japan)

Abstract: Sleep apnea seriously affects the quality of life and health. Polysomnography (PSG) is the "gold standard" for diagnosis, but it is expensive and inconvenient for long-term monitoring. This paper proposes a new method based on a universal smart wristband to detect sleep apnea. By analyzing the heart rate, blood oxygen saturation, and sleep state data collected by the wristband, an adaptive physiological data reconstruction method and a data interpolation method are used to achieve noise filtering. In feature engineering, continuous physiological variables and categorical variables are fused to deeply extract sleep state features. The classification module uses a lightweight gated recurrent unit model to simplify the training process and reduce the risk of overfitting. The experiment obtained 93.68% accuracy and 93.97% recall on a 23-person dataset. Correlation analysis find that blood oxygen saturation, body mass index, and age are confirmed as key features for determining sleep apnea. Compared with polysomnography, this method is more suitable for long-term monitoring in a home environment.

Key words: sleep apnea detection; universal wristband; multimodal data processing; long-term health monitoring; analysis of multi-factor impact indicators

收稿日期: 2024-09-02; 修回日期: 2024-11-14; 录用日期: 2024-11-19。

基金项目: 国家自然科学基金资助项目(62302145); 安徽省科技创新平台重大科技项目(202305a12020012)。

作者简介: 黄锦阳(1994—), 男, 安徽安庆人, 讲师, 博士, CCF会员, 主要研究方向: 多模态人因感知、人工智能; 崔丰麒(2001—), 男, 山东烟台人, 硕士研究生, CCF会员, 主要研究方向: 情感计算、人工智能; 马长秀(1980—), 女, 安徽蚌埠人, 主任医师, 博士, 主要研究方向: 睡眠呼吸暂停、肺疾病相关; 樊文东(1998—), 男, 山东德州人, 硕士, 主要研究方向: 多模态人因感知、机器学习; 李萌(1988—), 男, 安徽合肥人, 副研究员, 博士, 主要研究方向: 网络安全、车联网; 李经宇(1995—), 男, 安徽淮北人, 特任副研究员, 博士, 主要研究方向: 人工智能; 孙晓(1980—), 男, 山东烟台人, 教授, 博士, CCF会员, 主要研究方向: 多模态人因感知、自然语言处理、人脸识别; 黄林生(1977—), 男, 安徽安庆人, 教授, 博士, CCF会员, 主要研究方向: 图像处理、人工智能; 刘志(1986—), 男, 河北衡水人, 副教授, 博士, CCF会员, 主要研究方向: 多媒体信息处理。

0 引言

睡眠呼吸暂停是指睡眠过程中气道反复发生部分或完全塌陷的现象。根据美国国家健康研究院(National Institutes of Health, NIH)的数据,全球估计有近1亿人受到睡眠呼吸暂停的影响,其中阻塞性睡眠呼吸暂停(Obstructive Sleep Apnea, OSA)为最常见的形式。呼吸暂停不仅仅是睡眠质量的问题,它对个体的整体健康和日常生活产生了深远的负面影响^[1]。睡眠呼吸暂停会导致慢性睡眠剥夺和日间疲劳。反复罹患呼吸暂停会增加心血管疾病、高血压、糖尿病以及心脏病发作和中风的风险,影响生活质量。此外,患者还可能经历情绪波动、记忆力减退和工作效率下降,若不及时治疗甚至可能导致意外伤害和死亡风险增加^[2]。综合来看,呼吸暂停是一种普遍且影响深远的疾病,对呼吸暂停的准确检测具有重要意义^[3]。

在现代睡眠医学中,多导睡眠监测(Polysomnography, PSG)被广泛认为是诊断睡眠障碍的“金标准”。PSG通过监测睡眠期间的生理活动,如脑电波、眼动、肌电、心率以及呼吸模式,为诊断和评估各种睡眠障碍提供了详尽的数据^[4]。此技术对于确诊OSA以及制定个性化的治疗计划至关重要。然而,尽管PSG在诊断睡眠障碍方面具有无可比拟的优势,但其成本高昂、操作复杂,且需要专业人员的配合,这限制了它的普及和应用。病人在进行PSG监测时常常需要在医院度过整晚,连接多根导线,这不仅降低了舒适性,也影响了睡眠质量^[5]。此外,由于无法进行长期或远程监测,PSG在持续性病情跟踪和管理方面存在局限性^[6]。因此,尽管PSG是OSA的诊断的“金标准”,但在实际应用中,其局限性导致现有的OSA难以早发现、早诊断、早治疗^[7]。因此,研究和开发更便捷、经济且患者友好的睡眠呼吸暂停监测方法,对于改善OSA患者的诊断和治疗过程具有十分重要的意义。除了使用PSG设备进行睡眠呼吸暂停的检测,一些研究者考虑使用脉搏血氧仪和心电图(Electrocardiogram, ECG)数据来有效的检测阻塞性睡眠呼吸暂停^[8],但是该类方法中血氧数据和ECG数据的获取仍然需要专业的仪器,因此该类方法便捷性和普适性均较差,不利于患者在家中长期监测^[9]。

考虑到患者在呼吸暂停期间,其生理参数会发生显著变化,通过捕捉这一变化并进行相对应分析进而实现对呼吸暂停状态的捕捉也是行之有效的方案之一。具体来说,当发生呼吸暂停时,由于气流的中断或减少,患者的血氧饱和度会降低。这种降低可能导致组织缺氧,触发一系列生理反应。在缺氧的情况下,患者的心脏可能会加速跳动,以增加血液流动,将更多的氧气运送到身体各部分,从而使心率升高^[10]。这些变化可以被通用设备如手环所捕捉,进而为呼吸暂停的临床诊断和监测提供重要的依据。

目前大众普遍佩戴的通用运动手环可以采集部分睡眠生理数据,以帮助检测睡眠呼吸暂停。该设备较为便携,侵入感低,不会对患者造成负担。然而,通用运动手环采集的数据一般比较粗糙,精度不够,且目前市面上的运动手环仅支持健康检测和睡眠质量检测,均不支持呼吸暂停检测^[11]。因此,为了结合PSG和普通商用智能手环的优点,提高睡眠呼吸暂停监测的便捷性,进而实现对于呼吸暂停的长期监测,本文提出一种基于商用运动手环的睡眠呼吸暂停检测方法。该方法利用大众普遍佩戴的商用运动手环获取的精度相对粗糙的心率和血氧饱和度以及睡眠数据,在训练环节通过高精度PSG数据进行准确标注,进而根据睡眠阶段的不同实现睡眠数据的有效增强,并结合所提出的适合小规模睡眠数据的轻量级时序序列数据分类模型,最终实现在预测阶段仅利用手环数据即可对睡眠呼吸暂停进行准确检测的目标。本文的主要贡献如下:

1) 本文率先创建了真实医院场景下基于运动手环的睡眠监测数据集,其中包含了心率、血氧饱和度和睡眠状态的时序数据,这些数据都被高精度的PSG数据进行了同步标记,可以用于处理睡眠呼吸暂停监测的相关任务。

2) 针对不同睡眠时期生理数据的显著差异,本文设计一种基于异构时期特征的自适应生理数据重构方法,利用异构数据的时频域特征对生理数据所处的睡眠状态进行准确判断,并使用自适应的数据插值方法对手环所采集的生理数据进行有效重构,实现对噪声

和低频分量的自适应滤除,该方法在滤除无关分量的同时显著增强了收集信号中与呼吸行为相关的成分。

3) 针对睡眠数据集体量较小且睡眠呼吸暂停任务对于实时性的要求,本文基于门控循环单元(Gated Recurrent Unit, GRU)模型搭建了一个适配于智能手环所采集生理数据的呼吸暂停分类框架,并使用如随机森林(Random Forest, RF)、长短期记忆(Long Short Term Memory, LSTM)网络和深度信念网络(Deep Belief Networks, DBN)作为基准模型分别在数据集上进行性能评估,大量真实场景下的实验结果证明了基于GRU模型的分类框架能够更好地分类睡眠呼吸暂停状态,其中准确率能达到93.68%,召回率为93.97%。更为重要的是与LSTM和DBN等深度模型相比,提出的框架的训练速度平均提高了一倍。

4) 本文通过多种特征重要性分析方法来探究分类结果和数据特征之间的关系。经研究发现,血氧数据、身体质量指数(Body Mass Index, BMI)和年龄对睡眠状态分类结果有显著联系。不同的呼吸暂停次数在BMI、平均血氧、最小血氧、血氧标准差及心率标准差等特征上有着显著的组间差异,而在不同的睡眠阶段上无明显差异。类别指标如是否抽烟、是否患有糖尿病对呼吸暂停的分类也有着显著的影响。这些研究结果为探索睡眠呼吸暂停的原因提供了有效的数据支撑。

1 相关工作

1.1 基于机器学习的呼吸暂停检测方法

近年来,随着机器学习技术的快速发展和广泛应用,基于机器学习的睡眠呼吸暂停检测方法在医学和生物信息领域内得到了显著的关注和研究。不同于传统的睡眠监测方法,基于机器学习的方法提供了一种便捷、经济且患者友好的睡眠监测方案。通过使用特征提取、模式识别和多模态融合方法,研究人员能够从复杂的生理信号中提取有价值的信息,从而提高睡眠呼吸暂停检测的准确性和效率^[12]。这些方法旨在通过深入分析和处理生理信号,如氧饱和度(Oxygen Saturation, SPO2)^[13]、ECG^[14]、眼电图(Electrooculogram, EOG)^[15]或脑电图及脑地形图(Electroencephalogram, EEG)^[16]等,来自动识别和检测睡眠呼吸暂停的事件。

例如文献^[17]提出一种基于自适应多任务学习的睡眠生理时序分类方法。该方法利用单导脑电与心电检测睡眠分期作为标注,结合睡眠呼吸暂停低通气综合特征,综合设计自适应任务间关联性学习模块,利用通道注意力机制建模睡眠阶段和呼吸暂停低通气之间的相关性,最终实现对于呼吸暂停的准确。Furlotti等^[18]通过统计分析和机器学习方法探索神经活动变化以检测阻塞性睡眠呼吸暂停,使用了多种机器学习技术进行睡眠暂停的全夜预测。Tran等^[19]利用多种机器学习方法针对ECG缺失的频谱图在连续正压通气滴定期间提前检测睡眠呼吸暂停,该方法实现了对呼吸暂停的准确检测,其中在使用支持向量机(Support Vector Machine, SVM)分类器时的检测性能达到最佳,此时准确率为98.2%,F1分数为0.93。机器学习方法可以很好地处理心电图、眼电图、或脑电图等专用设备所收集的数据,并实现较高的呼吸暂停检测成功率,但由于通用手环采样率低,数据精准度差,传统的机器学习方法可能难以实现最佳性能,针对基于通用手环的数据处理方法还有待深入探索以实现高精度的睡眠呼吸暂停检测。

1.2 基于指环式和腕带式的便携式睡眠监测方法

便携睡眠监测设备通常采用呼吸、鼾声、血氧饱和度和心电信号 and 呼吸努力度等指标监测呼吸事件,其舒适方便、的优势加上较高的准确性,使之成为初筛OSA不可或缺的方法^[20]。根据穿戴方式的不同,现有的便携式睡眠监测方法主要分为指环式与腕带式。指环式睡眠呼吸监测仪通过内置的光传感器发射光线到皮肤下,利用光吸收的差异来测量血液中的氧气含量。当血液中的氧合血红蛋白与脱氧血红蛋白的比例发生变化时,光的吸收率也会相应变化,从而可以计算出血氧饱和度。

尽管指环式睡眠呼吸监测仪因其便携性和易用性而受到欢迎,

大多数指环式监测仪依赖于小型电池供电,这意味着用户需要定期充电,而频繁的充电需求可能会给用户带来不便^[21]。与指环式睡眠检测仪相比,腕表式睡眠监测仪有着更大的电池容量,因此续航问题得到了有效解决。腕表式睡眠监测仪是一种新型的睡眠监测工具,具有操作简单、无扰睡眠、连续测量等优点。此类设备多是通过探测末端指节血管容量的变化,来监测呼吸事件发生时的交感神经增强来判断呼吸事件及睡眠分期。以色列技术学院于2005年研制出的Watch-PAT腕表^[22],该腕表采用了一套独特方法划分睡眠分期,并通过患者的血氧饱和度、外周动脉张力、心率和活动记录仪来识别睡眠呼吸障碍,其研究表明Watch-PAT的检测方法与PSG高度一致,相关性可高达85%^[23]。但该研究结果并未被学者所广泛认同,BROWNJ等^[24]采用Watch-PAT对500例疑似OSA患者进行大规模监测,发现Watch-PAT的检测方法与标准PSG相比,有高达30%到50%的误诊率。这直接说明了该腕带式设备在呼吸暂停检测任务中并不普遍可靠。

本文通过表格方式展示了这些方法在数据采集上的差异,如表1所示。在采集设备及穿戴方式上,基于手环的方法是将手环佩戴于手腕处,类似于手表的形式,而基于指环和基于腕带的方式则分别把指环和腕带佩戴于手指上或手腕上,佩戴方式有差异点也有相同点。

表1 基于运动手环、指环以及腕带的睡眠呼吸暂停检测方法在数据采集上的差异对比

Tab. 1 Comparison of sleep apnea detection algorithms based on sports bracelets, finger rings and wristbands in data collection

不同方法	采集设备及穿戴方式	采集数据模态	采样率	供电容量及数据采集时间
基于运动手环	使用手环,佩戴于手腕处,类似于手表的形式	加速度信号、陀螺仪信号、光电容积脉搏波描记法PPG信号、GPS定位信号、温度信号、光敏传感信号	运动手环的加速度采样率:50~100 Hz,输出的心跳频率:1~5Hz	电池容量:100mAh~300mAh,数据采集时间:5~7天
基于指环	使用指环,直接套在手指上,便于携带且不易察觉	加速度信号、陀螺仪信号、光电容积脉搏波描记法PPG信号	指环的加速度采样率:20~50 Hz,输出的心跳频率:1~2Hz	电池容量:20mAh~50mAh,数据采集时间:1~3天
基于腕带	使用腕带,同样佩戴于手腕上,与运动手环类似,但更为简单,没有显示屏或者仅有简单的LED指示灯	加速度信号、陀螺仪信号、光电容积脉搏波描记法PPG信号	腕带的加速度采样率:20~50 Hz,输出的心跳频率:1Hz左右	电池容量:50mAh~100mAh,数据采集时间:3~7天

1.3 睡眠呼吸暂停开源数据集

现有的关于睡眠呼吸暂停检测数据集主要有以下三个,分别是侧重于睡眠阶段分类的Sleep-EDF数据库,侧重于呼吸暂停检测的Apnea-ECG数据集以及侧重于睡眠障碍评估的MASS数据集。

Sleep-EDF数据库包含两部分,第一部分是用于睡眠质量研究的多导睡眠图记录和睡眠评分,第二部分是研究睡眠障碍的年龄相关的记录,包括EEG、EOG、下颌肌电图(electromyogram, EMG)、呼吸努力和血氧饱和度等信号。此数据集主要用于睡眠阶段分类、睡眠质量研究、睡眠呼吸暂停检测等任务,并广泛应用于人类睡眠相关研究及方法的开发。Sleep-EDF数据库包含大量受试者的数据,数据量庞大,并且信号类型丰富,适用范围广,因此适合于机器学习和深度学习相关的睡眠检测方法的开发和验证。但是该数据集的某些模态数据质量相对欠佳,且部分数据集的信号类型有限,因此难以支持复杂的分析和研究。

与Sleep-EDF数据库的侧重任务不同,Apnea-ECG数据集^[27]是专门针对睡眠呼吸暂停研究的数据集,它收集了来自70多名受试者的夜间心电图(ECG)记录。每个记录均包括心电图信号和睡眠呼吸暂停事件的详细标注。该数据集提供了详细的睡眠呼吸暂停事件标注,主要用于开发和测试睡眠呼吸暂停自动检测方法,非常适用于研究ECG信号与睡眠呼吸暂停之间的关系。但该数据集的缺点较为明显,这主要体现其所收集信号集中于ECG上,缺乏其他类型的生理信号,因此其在多信号综合分析任务中的应用会受到一定的限制。

MASS(Montreal Archive of Sleep Studies)数据集^[28]收集了健康成人和睡眠障碍患者的多导睡眠图记录,包括EEG、EOG、EMG、ECG和血氧饱和度等信号。该数据集包含详细的睡眠阶段评分和一部分呼吸事件标注。MASS数据集提供了多种生理信号和详细的睡眠评分,适合于复杂的方法开发和睡眠研究,包括睡眠阶段分类、睡眠障碍诊

此外在采集数据模态上,基于手环的方法有着更多的数据采集模态,例如全球定位系统(Global Positioning System, GPS)定位信号、温度信号、光敏传感信号等,这些多模态数据均是基于指环方法和基于腕带方法所不具备的。在采样率上,运动手环的加速度采样率更高且输出计算的心跳值更频繁,这是因为运动手环相较于指环和腕带,它的电池容量更大,能支持的传感器性能也相应更高。在相对大容量电池的支持下基于运动手环的方法能持续长时间采集用户生理状态,一般单次充电能维持一周,因此适合文中的所述的睡眠呼吸暂停检测任务。

Chen等^[25]使用华为GT2智能手表与PSG系统进行关于OSA严重程度的分类准确度对比。与PSG相比,智能手表预测呼吸紊乱指数 ≥ 5 的OSA患者的准确性为81.1%,这说明了基于腕带设备在一定条件下能实现不错的呼吸暂停检测性能。不幸的是,尽管腕表式睡眠监测在诊断OSA方面较其他便携设备具有一定优势,但其较低的采样率使得当呼吸暂停连续发生时,系统的检测性能下降。此外边缘设备受限的计算资源也极大地影响了检测精度^[26],这些缺点限制了该类方法的广泛应用。基于此,有效的低采样数据增强策略以及轻量级的训练网络亟待被提出以适应通用手环的特点并实现对于呼吸暂停的高效准确检测。

断、睡眠呼吸暂停检测等。但由于该数据集覆盖范围过于广泛,导致部分特定类型的睡眠障碍数据较少,因此需要结合其他数据集综合使用以获得更全面的研究成果。

以上数据库各自提供了丰富的数据资源,对于利用机器学习或深度学习方法进行睡眠呼吸暂停的研究具有重要价值。然而,这些数据集的数据均需要专业的仪器和设备采集,难以在居家环境中进行测量,因此不利于睡眠呼吸暂停的居家长期监测。现有研究没有基于通用手环的呼吸暂停数据集,但通用手环由于其普适性及舒适性可以更好地适应于居家场景下呼吸暂停检测任务。为了弥补这一研究空白,本文构建了一个基于运动手环采集数据的睡眠呼吸暂停数据集,这是首个利用商用智能手环所采集的睡眠数据集,主要包含SPO₂、心率、睡眠状态和患者的各项生理信息,可以较好的用于睡眠呼吸暂停的检测研究。针对这一数据集,本研究也提出一种基于商用运动手环的睡眠呼吸暂停检测和分类方法,通过利用便携商用智能手环,来采集患者的血氧饱和度、心率和睡眠状态数据,实现对睡眠呼吸暂停的即时检测和准确分类。在采集数据的过程中时,患者只需要将手环佩戴至腕部,无需其他操作即可采集数据。该方法无需专业人士操作,步骤简单,且佩戴的手环不会带来任何不适感,因此并不会影响患者的睡眠质量。患者可以很方便地佩戴通用手环在家中监测,便于睡眠呼吸暂停的长期监测和研究。

2 睡眠呼吸暂停检测框架及具体技术

本文基于运动手环采集到的生理信号制作数据集,利用数据预处理以及数据增强方法滤除信号中与睡眠呼吸暂停行为无关的分量与噪声,并通过针对性的特征工程以及多种分类模型进行分类检测,以实现评估模型性能的同时输出分类结果的目标。最后通过分析特征的重要性,判定具体哪些身体类别指标与睡眠呼吸暂停高度相关。

睡眠呼吸暂停检测流程如图1所示,主要由数据采集及预处理、数据增强、特征工程、分类模块、特征重要性分析等五个模块所构成。

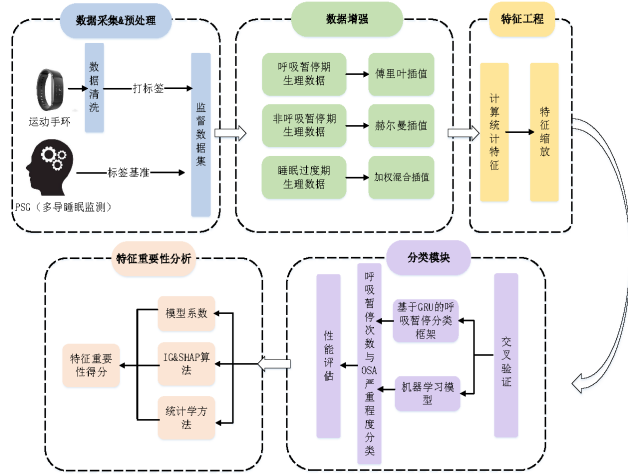


图1 基于通用手环的睡眠呼吸暂停分类检测框架

Fig. 1 Sleep apnea classification detection framework based on a universal wristband

2.1 预处理、数据增强及特征工程模块

在数据采集阶段,本文使用运动手环和PSG设备分别采集受试者的睡眠生理数据。运动手环采集到包括心率、血氧和睡眠阶段数据,这三种时序数据被整合到一起,以便于后续分析和处理。随后对采集到的数据进行预处理工作,并由有经验和专业知识的医生参照PSG的基准数据进行人工数据标注,得到监督数据集。

在数据增强阶段,针对不同睡眠时期生理数据的显著差异,本文设计一种基于异构时期特征的自适应生理数据重构方法,利用异构数据的时频域特征对生理数据所处的睡眠状态进行准确判断,并使用自适应的数据插值方法对手环所采集的生理数据进行有效重构,以实现噪声和高频分量的自适应滤除。其中时期特征的定义为:不同睡眠时期下的生理数据的特征。方法的流程如图2所示,包括:

S1: 对手环数据集 D_1 进行预处理,同时使用高斯低通滤波器对数据进行滤波,滤除高频分量和噪声,最终得到数据集 D_2 。

S2: 定义滑动窗口 n 。该滑动窗口内容纳一定时间内所有采样点的数据序列,用于在手环采集的时序生理信号上进行滑动,分批处理生理数据,对窗口内的数据进行插值和数据重构。

S3: 使用短时傅里叶变换 (Short-Time Fourier Transform, STFT) 以及常用统计量 (方差、偏度、峰度) 来对窗口内数据的特征进行提取,通过对比中低频分量和统计特征这些计算指标来判断窗口内数据的睡眠状态 $state$ (是否发生呼吸暂停,是否在过渡期间),进而自适应的调整数据插值方法和权重大小。

S4: 判断 $state$, 并根据生理数据所处的不同睡眠状态来自适应的对生理数据进行重构。

S5: 经过自适应数据重构后,得到数据集 D_3 。

S6: 对重构过的数据集 D_3 进行自适应滤波得到数据集 D_4 。自适应滤波器可以实时调整滤波器参数,以适应输入信号的变化,进而滤除在数据重构期间产生的噪声和高频分量,保证数据的准确性。

在特征工程模块,与仅考虑生理数据方法不同,本文融合传感器所获取的连续生理变量以及身体先验状态信息的类别变量,综合实现了对睡眠状态特征的深度提取。本文先对智能手环采集的数据进行预处理以实现对于不同模式数据的有机融合。考虑到制作的数据集既包含心率、血氧饱和度等连续变量,又包含是否抽烟喝酒,是否患有高血压糖尿病等类别变量,而不同特征的数据维度均不在一个等级上,因此本文拟对数据型变量进行归一化和标准化,使不同变量位于相同的尺度。归因于机器学习方法对特征的尺度较为敏感,这一操作显著提升了机器学习模型分类精度。而对于深度学习模型,归一化和标准化也有助于加快模型训练速度和提高模型稳定性。

在完成数据的归一化及标准化后,系统针对不同类别变量分别进行缺失值和异常值进行处理。对于连续变量的缺失值和异常值,系统使用中位数来填补空缺。而对于类别变量,系统拟采用频率最高的值作为细节修正。

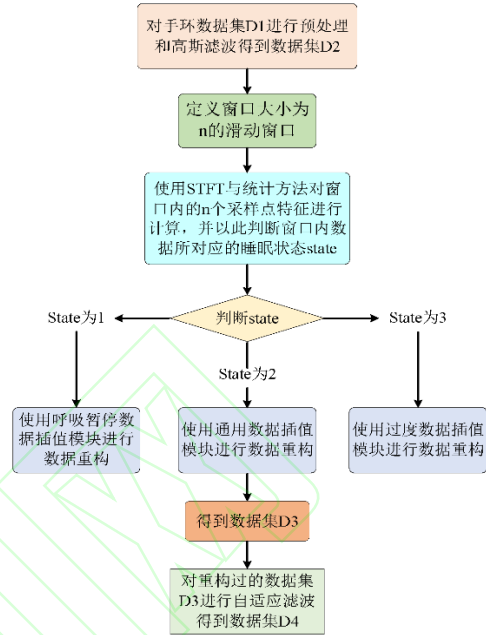


图2 基于时期特征的自适应插值方法流程

Fig. 2 Process of adaptive interpolation algorithm based on temporal features

在特征提取阶段,本文将从手环采集的时序数据中提取用于训练的特征集。由于商用智能手环与专门用于临床检测的医疗设备不同,其采集的数据精度相对粗糙,不够精确。因此,本研究仅从手环采集的血氧饱和度和心率生理信号中提取了部分统计特征,包括平均值、标准差、最大值、最小值,而没有其计算频域特征(比如,功率谱密度等)或时序特征,这很好地避免了由粗糙数据以及采样率低带来的特征失真问题。除了这些通用特征,考虑到呼吸暂停行为发生之后,人体内的血氧水平会显著降低,本文还利用血氧饱和度特征计算出了平均血氧降低水平(Average Oxygen Desaturation Level, AODL)这一特征,该特征能够。具体来说,平均血氧降低水平是通过监测受试者在一段时间内的连续的血氧饱和度数据来计算的,基本思路是确定在一定时间内血氧饱和度的平均水平的降低幅度。计算AODL时,首先要记录每次测量得到的血氧饱和度数值,进而计算不同采集次数过程中血氧饱和度的平均降低水平。通过确定每次测量值与个体正常水平的差值,随后计算这些差值的平均数,实现对AODL特征的有效计算。在完成特征提取之后,系统开始进行特征构造。由于数据集中含有类别变量,比如睡眠阶段,该变量需要使用字符来更好地表示不同的睡眠阶段。基于此,系统对类别变量进行编码,以便模型能更好地处理数据。在完成特征构造后,系统对处理后的数据集进行特征工程最后阶段的特征融合。通过对手环采集的血氧和心率数据以及睡眠阶段这三种时序数据整合到一起,并结合病人健康信息相关的类别变量(例如,是否患有某项疾病、年龄、身高、体重等),系统完成了所有的特征提取,针对每个滑动窗口样本提取了共计21个统计或类别特征。具体包括:心率模态的平均值、最大值、最小值、标准差;血氧饱和度模态的平均值、最大值、最小值、标准差、平均血氧降低水平;以及睡眠阶段、年龄、身高、体重、体脂率、吸烟习惯、饮酒习惯、心脏疾病史、高血压史、糖尿病史、抑郁或焦虑症状的心理健康评估、BMI(对应图6图7以及图8的特征)。

2.2 分类模块

准确的睡眠呼吸暂停检测是评估受试者睡眠质量和身体健康状况的基础。本文旨在利用便携式呼吸暂停检测设备,为用户提供事

后睡眠健康的预警服务。具体而言,该设备将生成详细的睡眠质量报告,并对可能存在的睡眠呼吸障碍等健康问题提供及时、准确的预警。基于此设计合适的分类模块是实现文章最终目标的重要基石。

在分类模块,考虑到呼吸暂停任务对于实时检测的迫切性以及手环等边缘设备计算资源的有限性,系统设计了一种基于门控循环单元的轻量级时序序列数据分类模型,与长短期记忆(LSTM)人工神经网络相比,GRU模型通过将遗忘门和输入门合并为一个单一的更新门,减少门控数量,极大地减少了模型的参数量,进而使得模型相对更容易训练,实现轻量化的目标的同时降低了过拟合的风险,另一方面与深度信念网络(Deep Belief Networks, DBN)的训练模式相比,GRU不需要复杂的预训练过程,可以直接采用有监督或无监督的方式进行端到端的训练,这大大提高了训练效率,这些因素最终实现对于呼吸暂停的高效检测。

本研究所使用的数据集是由21名试验者的睡眠数据构成,数据集的规模相对有限。因此,本文工作并不适于使用参数众多、结构复杂的模型来进行分类研究,后者的模型训练往往需要更大的数据量来支撑。基于此,本文使用GRU模型为基础,通过联结全连接层和Dropout层,最后使用Softmax层作为输出,搭建了一个轻量级的睡眠呼吸暂停检测和分类框架,并对比随机森林(Random Forests, RF)、LSTM、DBN、K近邻方法(K-Nearest Neighbor, KNN)和多元线性回归(Mixed Logistic Regression, MLR)模型在数据集上进行训练和测试,最后进行性能的分别评估,实现不同模型的性能差异的有效评估。系统使用交叉验证技术来测试模型性能并选取最佳超参数。具体通过使用自定义的参数网格来容纳要测试的所有参数组合,然后以分类准确率、召回率和F1分数的组合作为评分标准,分别获得各个模型的最佳参数集合,并得到模型在该参数下的平均性能指标。在经过性能评估之后,综合选出性能最好的模型,并结合该模型的最佳参数组合,再使用最佳参数在整个训练集上重新训练模型以实现在测试集上的最终评估,最后获得模型的综合性能指标。

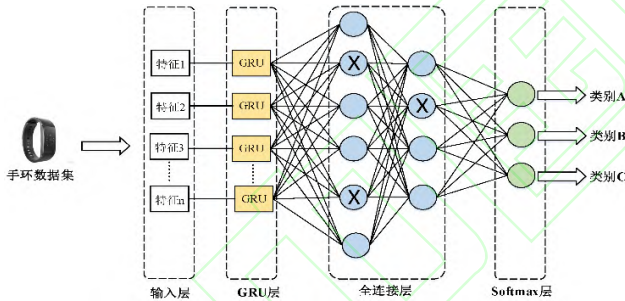


图3 基于门控循环单元的轻量级时序序列数据分类模型

Fig. 3 A lightweight temporal sequence data classification model based on gated recurrent units

图3展示了本文提出的基于门控循环单元的轻量级时序序列数据分类模型。本文使用GRU模型来搭建深度学习分类框架。GRU是一种高效的循环神经网络架构,专门为解决标准循环神经网络(Recurrent Neural Network, RNN)在处理长序列数据时面临的梯度消失问题而设计。与其它RNN变体如长短期记忆网络LSTM相比,GRU模型结构更简单,参数更少,识别效率更高,因此更加适合本研究对于实时性的要求,且在大多情况下GRU模型能够提供与LSTM类似的性能。

本文使用通用智能手环所采集的精度相对粗糙的生理数据作为输入,结合PSG数据作为基准,将处理后的数据输入GRU层进行训练。通过在GRU层后面接入全连接层,以进行特征的进一步组合和转换。考虑到睡眠呼吸暂停数据集相对较小,为了防止过拟合现象的发生,系统在在全连接层中添加Dropout层。最后,系统使用Softmax层作为输出层,输出模型对每个呼吸暂停任务中每个类别的预测概率。对比方法多元线性回归MLR是一种用于处理多类别分类问题的统计模型,是二分类逻辑回归的推广方法。模型不仅可以预测样本属于哪个类别,还能给出样本属于各个类别的概率,以帮助了

解睡眠呼吸暂停的严重程度。MLR的核心是Softmax函数,它将线性函数的输出映射到概率分布上。在有 m 个特征, n 个类别的分类问题中,模型的目标是估计输入特征属于每个类别的概率。模型方法如式(1)所示:

$$P(Y = j|X = x) = \frac{e^{\beta_j + \beta_j^T x}}{\sum_{k=1}^n e^{\beta_k + \beta_k^T x}} \quad (1)$$

P 是类别 j 的预测概率, β_{j0} 是类别 j 的截距项, $\beta_j^T x$ 是与类别 j 相关的特征权重向量,而 $\beta_j^T x$ 是特征向量 x 与权重 β_j 的点积。在睡眠呼吸暂停的多分类任务中,本文收集患者的生理数据作为特征,并标注其对应的呼吸暂停严重程度作为目标变量。然后,使用MLR模型进行训练。模型会学习每个严重程度级别对应的权重 β 和截距项 β_0 ,从而最终输出该患者在每个呼吸暂停严重程度类别上所对应的预测概率。

另一种对比方法KNN是一种广泛应用的监督学习方法,其核心思想是基于近邻的概念,即一个样本的预测结果主要由其最近邻的样本的已知结果决定。当给定一个未标记的样本,该方法会在训练集中寻找与该样本最接近的 K 个已标记样本,然后根据这些近邻样本的信息来预测当前样本的标签,这种方法无需训练,能快速实现类别分类,因此在一定程度上适配睡眠呼吸暂停检测任务。KNN方法的具体原理如式(2)所示:

$$\hat{y} = \arg \max_c \sum_{i=1}^K I(y_i = c) \quad (2)$$

其中, \hat{y} 是预测的类别, $I(y_i = c)$ 是一个指示函数,当第 i 个邻居的类别等于 c 时取值为1,否则为0。此函数计算的是在最近的 K 个邻居中类别属于 c 的样本数。而预测结果则是这些数目中最大数所对应的类别。

因此,为了合理运用KNN作为对比方法,本文通过转换数据获得标准的距离度量,以解决混合型数据的问题。首先对连续变量进行标准化。对于连续变量,使用标准化的方式来处理,确保各个特征在计算距离时具有相同的尺度。具体来说,要将每类特征值减去该特征的均值并除以标准差,进而使得每个连续特征的均值为0,标准差为1。而对于类别变量,考虑到独热编码会为每个类别的每个可能值创建一个新的二进制特征,表示该值是否出现,本文使用独热编码(One-Hot Encoding)将其转换为数值型格式。在对连续变量和类别变量分别进行处理之后,可以将两者组合成一个新的特征矩阵,以实现特征融合,并使用欧氏距离来计算其距离。由于类别变量已经被转换为二进制形式,欧氏距离在这些变量上的计算实质上已经等同于使用汉明距离。在睡眠呼吸暂停的多分类任务中,KNN分类器有着较好的适配性,这是因为KNN方法可以根据患者的多个生理指标来预测他们的呼吸暂停等级,探究患者受呼吸暂停疾病的影响程度。

另一种机器学习对比方法RF是一种集成学习方法,主要用于分类和回归任务,通过构建多个决策树(Decision Trees, DT)并融合它们单独的预测结果实现对睡眠呼吸暂停情况的分类。由于睡眠呼吸暂停分类是一个多样性和多源性并存的复杂任务,这意味着相同的呼吸波形对应到不同人身上可能对应的呼吸暂停严重程度并不一样,这就要求分类模型有着较好的多样性。对于呼吸暂停检测任务,对比方法RF的引入可以在提升模型多样性的同时避免过拟合,并有效提升估计精度。

2.3 特征重要性分析模块

2.3.1 机器学习特征重要性分析方法

在基于通用手环的睡眠呼吸暂停检测任务中,分析不同特征的重要性具有多方面的意义。这不仅有助于提高模型的准确性和鲁棒性,还能为临床诊断提供有价值的参考信息。在数据特征维度,通过识别哪些数据特征对预测结果影响最大,可以针对性地优化方法,减少计算资源消耗,同时提升模型的预测准确性。而在人的自身属性维度,如是否肥胖、是否有糖尿病,收集和分析大量人群的特征数据,有助于发现某些特定群体中呼吸暂停发病率较高的趋势,为制定公共健康政策提供科学依据。

每个特征重要性分析模块是对呼吸暂停分类结果和类别变量以

及连续变量的之间的潜在关系进行深入探究,通过多种特征重要性分析方法输出特征的重要性评分,以此探索不同变量与呼吸暂停之间关系的密切程度以及呼吸暂停的主要影响因素,为探索呼吸暂停的成因提供了充分的理论分析与实验证明。分类模块和特征重要性分析模块作为本文实验框架的核心部分,共同构成了睡眠呼吸暂停检测系统。

因为本文所提出的方法以及对比的机器学习模型均带有特征重要性评估的功能,所以对于本研究中的多分类任务,可以直接使用上述方法来探究分类结果与特征之间的关系。以上机器学习方法都可以提供有关特征如何影响分类决策的信息。

对于MLR方法而言,其使用coef_系数数组来反映特征对分类结果的影响,这是因为该方法通过估计每个类别相对于参考类别的对数几率来进行分类。特征的系数大小和符号表明了该特征对分类结果的影响程度和方向。系数的绝对值越大,特征对分类结果的影响越大。正系数表明随着特征值的增加,某类的对数几率增加,即该特征的增加会增加属于该类别的概率,而负系数则表示对数几率减少,表示负相关。通过coef_系数的计算可以清晰的获取不同特征对于呼吸暂停分类的重要程度及其关联情况。

对比方法RF和KNN输出的是整体特征重要性,而非针对单个分类结果。其中RF衡量的是特征对模型预测能力的整体贡献,而不区分不同的分类结果。在RF中,特征重要性通常是通过观察当从决策树中移除某个特征时,模型性能的减少程度来确定的,可通过特征的平均不纯度减少来衡量。平均不纯度减少是用来衡量每个特征分裂点对模型预测能力的贡献的,即特征对分裂节点纯度的贡献总和的平均值。因此对比方法RF是通过平均不纯度减少来度量不同变量特征对于呼吸暂停分类的重要程度。

另一种对比方法KNN其本身不提供直接的特征重要性评估方法,但可以通过模型不透明性技术,比如置换重要性来间接评估特征的影响。置换重要性是通过随机打乱每个特征的值并观察模型性能的变化来衡量特征的重要性。本文通过逐个移除特征并评估模型性能变化的方法来判断该特征的重要性,如果移除某个特征后模型性能显著下降,则可以认为该特征对模型来说是重要的。性能下降越多,特征越重要。通过该方法,不同变量特征在KNN分类方法中对于呼吸暂停分类的影响情况被有效探索。

2.3.2 深度学习特征重要性分析方法

系统通过集成梯度法(Integrated Gradients, IG)来进一步探究所提出的深度学习模型中特征与分类结果之间的关系。IG是一种用于解释深度神经网络决策的技术,特别适用于揭示输入特征对模型预测结果的贡献度。它基于模型的梯度,即模型输出相对于输入的导数,来表示输入变化时输出的变化率。IG方法的核心思想是计算从一个基线输入到实际输入之间的路径上的梯度积分。基线代表数据的中性状态。这条路径上的积分表示了实际输入相较于基线在模型预测中的贡献度。方法公式如式(3)所示:

$$IG_i(x) = (x_i - x'_i) \times \int_{\alpha=0}^1 \frac{\partial F(x' + \alpha \times (x - x'))}{\partial x_i} d\alpha \quad (3)$$

其中: x_i 是输入 x 中第 i 个特征的值, x'_i 是基线 x' 中第 i 个特征的值。 $\frac{\partial F(x)}{\partial x_i}$ 是 $F(x)$ 相对于 x_i 的偏导数,表示 x_i 的微小变化如何影响 $F(x)$ 的值。积分 $\int_{\alpha=0}^1$ 沿着从基线 x' 到实际输入 x 的路径计算,表示对整条路径上的梯度进行积分。

在睡眠呼吸暂停检测任务中,首先,需要计算基线向量 x' 。基线向量 x' 代表着无呼吸暂停的正常状态,其对于连续变量和类别变量有着差异的表现形式,对于连续变量为特征,其基线向量为最低值或零值;而对于类别变量,其基线向量则表示为“无”或“正常”状态的类别。获取基线向量后,便进行集成梯度的计算。具体来说,对于每个特征,计算从基线到实际输入的集成梯度,进而获取每个特征对模型预测中实际输入属于各个呼吸暂停类别的概率有多少贡献。最后,通过分析各个特征的集成梯度值便可识别哪些特征对模型做出特定

分类起到至关重要的作用。

2.3.3 统计分析方法

尽管本文已经介绍了不同对比机器学习方法所对应的特征重要性分析方法以及提出的深度学习方法所对应的特征重要性分析方法,但这两类方法更多的是针对模型本身进行分析和探究,而忽视了数据集中数据分布可能存在的有效信息。因此,为了提供更加全面的分析以探究睡眠呼吸暂停的真正成因,本文将使用统计分析方法对数据集的自身分布进行准确测算。

卡方检验(Chi-squared Test)是一种基于观察值与期望值之间差异的统计方法,其基本思想是比较实际观察的频数与在两个变量独立的情况下所期望的频数之间的差异,若差异越大,则表明两个变量不独立的可能性越高,计算公式如式(4)所示:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (4)$$

其中: Σ 代表求和遍历所有单元格, O_{ij} 表示观察到的频数,而 E_{ij} 则表示期望的频数。卡方检验不仅可以检验两个类别变量之间是否独立,同时可以检验一个类别变量的观察频数分布是否符合某个特定的期望频数分布。而在呼吸暂停检测任务中,卡方检验主要用于特征选择,帮助确定哪些特征与目标变量之间存在显著的关联,从而评估一个类别变量的特征与呼吸暂停分类结果之间的隐含关系,最终为特征选择提供依据。

具体的对于每个特征,根据该特征的类别和目标变量的类别构建列联表。然后针对每个特征,使用上述公式计算卡方值,进而评估该特征的观察频数分布与在目标变量类别完全独立时的期望频数分布之间的差异。在评估完两者之间的差异性之后,根据计算出的卡方值和相应的自由度,进而查找卡方分布表以确定P值,从而实现对于差异显著性的评估。最终,根据高卡方值的特征与目标变量相关性更强,在构建分类模型时可能更有用这一基本原则,基于卡方值来选择与目标变量有显著关联的特征,以实现对于呼吸暂停分类任务的高效特征筛选。

由于卡方检验只能用于探究类别变量特征的重要性,而无法实现对于连续变量重要性的有效度量。因此针对于数据集中的连续变量,本研究拟采用方差分析(Analysis of Variance, ANOVA)的方法实现特征重要性的准确度量。ANOVA可以在多分类任务中用来分析不同分类结果之间的连续变量是否存在显著差异,进而提供特征如何影响分类结果的信息。ANOVA的基本思想是比较组内变异和组间变异。其中组内变异是指同一组内不同数据点之间的差异,而组间变异是不同组之间数据均值的差异。具体来说,ANOVA使用F统计量来测试各组均值是否相等,可通过式5进行表达:

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}} \quad (5)$$

其中: MS_{between} 是组间平均平方,反映不同组之间均值的差异,而 MS_{within} 是组内平均平方,其反映同一组内各数据点围绕组均值的波动。如果计算出的F值大于临界值,则可以拒绝原假设,即认为至少有两个组在血氧水平上的均值存在显著差异。在本研究中,使用ANOVA可以确定不同级别的呼吸暂停在某些关键生理指标上是否存在显著的均值差异,进而衡量不同程度的呼吸暂停在血氧饱和度、心率等连续变量上的差异性。若F相对于临界值足够大,则表明至少存在一次呼吸暂停的严重程度在这些指标上与其他次数呼吸暂停的严重程度有显著差异。

进一步的,本文使用多重比较方法来确定哪些组之间的差异是显著的。由于Tukey's Honestly Significant Difference (HSD)方法能够有效地控制多重比较中的一类错误,即假阳性错误,系统使用Tukey HSD方法来比较所有组之间的均值差异,其计算公式如式(6)所示:

$$q = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{\frac{MS_{\text{within}}}{n}}} \quad (6)$$

其中: \bar{x}_i 和 \bar{x}_j 分别代表两个不同组的样本均值, n 为每组的样本数, MS_{within} 为组内平均平方。

在本研究任务中,通过多重比较实现细化方差检验的目标,进而探究出某一个连续变量在不同的分类结果之间是否存在显著差异。本文旨在结合医学背景和上下文知识,更深层次探究连续变量特征与分类结果的关系。通过更进一步的研究,分析出呼吸暂停的严重程度是如何具体地影响不同生理指标,以最终实现为诊断和治疗提供更精确信息的目标。

3 实验结果与分析

3.1 数据集

3.1.1 数据采集

在本项研究中,共计采集了21名受试者的睡眠数据,累计达约174.4小时。其中,15名受试者经医学诊断确认为睡眠呼吸暂停综合症患者,余下6名受试者作为健康对照组参与研究工作。受试者年龄均匀分布在25到70岁之间。

为了确保环境的一致性和数据的准确性,如图4所示,所有数据的采集工作均在医院设置的睡眠监测室内进行。受试者在整个睡眠期间佩戴了智能手环和配置有多导睡眠仪(PSG)设备的数据采集线,以同步记录睡眠过程中的生理变化。本研究采用了商用智能手环采集的数据进行睡眠呼吸暂停的检测研究。



图4 医院监护室真实场景下数据采集

Fig. 4 Data collection in a real scene of hospital Intensive care unit

收集数据所构建的数据集其特点是包含手环中的三个异构模式数据,分别是血氧饱和度、心率、以及睡眠阶段信息,以及对应的多导仪PSG的标注。一个新数据集的样例如表2

所示:

表2 数据集中数据案例

Tab. 2 Dataset Data Example

start_time	end_time	SpO2	BPM	sleep_stage	PSG
1705413120	1705413180	95	72	a	0
1705413240	1705413300	94	75	b	0
1705413360	1705413420	92	78	a	1
1705413480	1705413540	90	82	c	2
1705413600	1705413660	94	74	a	0

其中start_time是开始时间戳,end_time是结束时间戳,两者间隔60秒,相邻帧也间隔60秒。SpO2代表这段时间内的平均血氧,BPM代表这段时间内的平均呼吸频率,sleep_stage则代表了不同的睡眠阶段,具体的字符“a”代表浅睡眠阶段,“b”代表深睡眠阶段,以及“c”代表快速眼动睡眠阶段,而最后的PSG代表了这段时间内呼吸暂停发生的次数。

除了这些采集自手环的连续变量信息,数据集中还记录了例如是否肥胖、是否饮酒、是否有糖尿病、高血压病史、心脏病史、年龄、以及BMI等类别信息,以期了解这些类别变量与睡眠呼吸暂停之间的关联性。

血氧饱和度和心率以及睡眠阶段信息均是时序序列,对应的采样率分别是每120秒一次、每30秒一次和每60秒一次,采样率设置较低的原因是因为本文方法并不是通过短期数据对一个受试者的睡眠

呼吸暂停严重情况进行判定,而是对每个受试者采集八小时的睡眠数据,以期通过长时段的监测来更综合地判定受试者的睡眠呼吸暂停的严重程度。所有受试者的实验样本经切割后达到8200条。

为了将手环采集的数据导入到本地,然后将这些数据与PSG设备的记录进行了细致对比,本文为此开发了一个APP,APP的界面如图5所示,具体包括心率、血氧数据的保存以及处理与检测。

值得注意的是,本文工作在医院里进行数据集的采集与构建是因为本文方法需要使用多导仪PSG数据作为基准真实值(ground truth)来给手环数据进行标签,因此才可以获得两者的相对误差。PSG多导仪的标注是通过手动完成的,标注了这段时间(滑窗时间内)是否发生了呼吸暂停,以及对应的呼吸暂停的次数。具体的,本文方法需要在训练的过程中使用多导仪PSG数据来提供标注,但一旦模型训练完成之后,训练好的模型可以应用于任何场景下的分类,即分类的过程中无需使用多导仪PSG数据,仅需要获取的手环数据即可,因此适合居家场景。

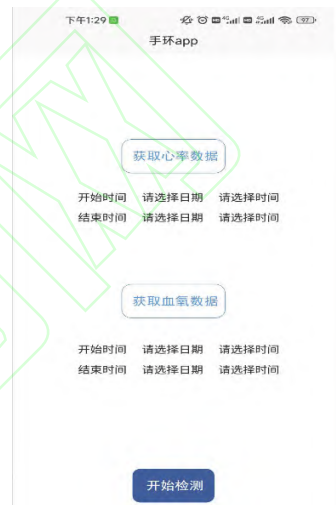


图5 手环APP实测采集界面

Fig. 5 Measurement and collection interface of the Bracelet APP

3.1.2 数据预处理

在数据处理阶段,本文采用三分钟的时间窗口作为数据切割的基准,对每个时间段内记录到的呼吸暂停事件进行了细致的分析和标注。标注过程中,每个时间段的呼吸暂停事件被分类为五个等级:0、1、2、3、4,以量化呼吸暂停的严重程度,并提供后续的数据分析与模型训练。

本研究使用商用智能手环采集血氧饱和度、心率和睡眠阶段等生理数据。鉴于商用智能手环的时间精度相对较低,并非专为监测睡眠设计,因此本研究只对数据的统计特征进行了提取。具体而言,对于心率和血氧饱和度数据,本文分别提取了四种基本的统计特征:平均值、最大值、最小值以及标准差。而对于血氧饱和度数据,本文还额外提取了平均血氧降低水平这一特征。平均血氧降低水平指的是在监测期间内,个体血氧饱和度降低的平均程度,这一指标可以反映血氧饱和度降低的频率与程度,对于评估睡眠质量及呼吸健康状况具有重要意义。在睡眠阶段数据的处理方面,智能手环采集的睡眠数据通常主要含在三个阶段,分别是浅睡眠、深睡眠,以及快速眼动睡眠阶段。因此,为了便于数据分析与处理,本研究采用了符号标注法,即使用字符“a”代表浅睡眠阶段,“b”代表深睡眠阶段,并使用“c”代表快速眼动睡眠阶段。除了上述从智能手环获得的生理参数外,本研究还通过临床调研,收集了参与者的个人信息及相关健康数据,包括年龄、身高、体重、体脂率、吸烟与饮酒习惯,以及是否存在心脏疾病、高血压、糖尿病等可能诱发睡眠呼吸暂停的基础疾病。考虑到心理因素可能会影响呼吸暂停的严重程度,本研究还对参与者进行了心理健康评估,以判断其是否存在抑郁或焦虑症状。

综合这些生理参数、个人信息及健康数据,本研究构建了一个全

面的数据集,为后续的睡眠呼吸暂停检测研究奠定了基础。通过构建此多模态数据集,本文旨在探索使用非专业设备在自然睡眠环境中进行睡眠呼吸暂停检测的可行性,同时评估多种因素对睡眠呼吸暂停诊断的影响,以期提高检测的准确性和实用性。由于真实采集场景中,用户经常会有异常行为,例如取下手环,这导致数据采集过程中存在部分数据缺失,因此本文对数据集实施了数据清洗。对于连续变量,系统采用中位数来填补其缺失值,以避免极端值对整体分布的影响。而对于类别变量,则选择频率最高的类别作为其填补的缺失值。

3.2 实验方案

本研究涉及的数据集包含了连续变量及类别变量,因此需要采取合适的处理方法以保证数据分析的有效性。本研究通过对类别变量应用独热编码(One-Hot Encoding)技术,将其转换为机器学习模型能够有效处理的数值型变量。独热编码通过为每个类别创建独立的二元特征,可以避免模型对类别数值大小的误解,同时可以提升模型的预测性能。

由于本数据集中包含的连续变量与类别变量在数值范围和分布特性上存在显著差异,本研究使用归一化处理策略,以确保不同类型的数据特征在后续分析中能够被公平地比较和评估,优化数据分析的有效性。通过 Shapiro-Wilk 测试对数据集的连续变量进行正态分布假设检验,结果显示这些变量均不服从正态分布。因此,本文选择使用最小-最大缩放(Min-Max Scaling)方法对特征进行归一化,计算公式如式7所示:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (7)$$

其中, X_{\min} 和 X_{\max} 分别为数据在特定特征上的最小值和最大值。通过该方法,所有特征的最小值变为0,最大值变为1。最小-最大归一化方法将原始数据 X 转换为缩放后的数据 X' ,并将所有特征值调整至相同的数值范围内,以优化方法的学习效率并提升模型的整体性能。

本文使用基于GRU的深度学习分类框架以及等多种对比基线模型在数据集上分别进行训练,并通过准确率、召回率、F1分数等性能指标评估模型在睡眠呼吸暂停数据集上的表现。进一步的基于模型预测的结果对患者夜间睡眠期间呼吸暂停次数进行计算以实现对患者呼吸暂停的严重程度进行分级,包括正常、轻度、中度和重度分类。

本研究使用交叉验证技术来获取模型的最佳超参数组合,并在此参数下测试模型的性能。具体的,系统分别确定每个模型需要调整的超参数,并自定义一个参数网格,其中包括想要测试的模型的所有参数组合。随后通过使用准确率、精确率、召回率和F1分数来自定义评分标准字典,以此评估模型性能,其中字典的键是评分标准的名称。本文使用 GridSearchCV 方法来传递评分标准的字典,在 scoring 参数中指定多个评分标准,进而使模型对每一个评分标准分别进行优化。

为了避免过拟合现象的发生,本文使用五折交叉验证,将数据集分成5份,其中4份用于训练,剩下1份用于验证。通过重复5次该训练过程,每次选择不同的验证集,并对所有结果进行平均实现对于模型性能的准确评价。GridSearchCV 方法通过尝试参数网格中的每一种参数组合,并使用交叉验证来评估每组参数的模型性能,来搜索最佳参数。随后,本文通过各个模型在最佳参数组合下的平均性能指标,经过性能评估,选出综合性能表现最好的模型。最后,将数据集按照7:3的比例分为训练集和测试集,并安排该模型在其最佳参数组合设定条件下重新进行训练,进而评估不同模型的最终性能。

本文拟基于模型测试的结果对夜间呼吸暂停次数进行计算,进而对患者的呼吸暂停严重程度进行分级。具体而言,模型对每条数据的每种分类结果进行概率预测,通过计算预测概率与其对应标签值的乘积求得的数学期望,得出在特定时间段内发生的呼吸暂停次数。例如,若模型预测某数据为分类“0”的概率为10%,为“1”的概率为30%,以此类推,那么该数据的数学期望即为该时间段内发生的预期呼吸暂停次数。通过对所有睡眠数据的数学期望求和,得到夜间睡眠期间呼吸暂停的预测总次数。最终,将这一预测值与多导睡眠图(PSG)数据报告的诊断结果进行对比,验证机器学习模型在呼吸暂

停严重程度分级上的准确性。

3.3 实验结果分析

本节比较了所提出的基于GRU的呼吸暂停分类模型与多种基线机器学习和深度方法在数据集上进行交叉验证的实验结果,实验结果如表3所示,其展示了模型的平均性能以及不同模型所需的训练时间,所有实验是在单卡4090的计算设备上进行。

表3 不同模型呼吸暂停分类结果对比

Tab. 3 Comparison of classification results of different models for respiratory pauses

方法	Accuracy/(%)	Precision/(%)	Recall/(%)	F1	训练时间/min
MLR	75.38	81.37	71.02	0.68	3.9
KNN	84.73	82.92	77.59	0.78	4.1
RF	92.21	90.65	89.63	0.90	3.6
GRU	93.68	94.36	93.97	0.92	5.4
LSTM	90.33	89.21	90.54	0.88	13.4
DBN	87.43	88.12	87.44	0.90	15.2

经实验结果对比,所提出的基于GRU模型的呼吸暂停分类方法在数据集上获得了优异的性能表现,其中准确率达到93.68%,精确率为94.36%,召回率为93.97%,F1分数为92.89%,各项指标都显著优于对比方法。更重要的是,在训练时间上,相比于使用多个门参数的LSTM以及需要预处理的DBN,所提出的轻量级网络能显著降低训练时间,节约一倍以上训练时间。与传统的机器学习方法相比,提出的方法训练时间稍高,但两者基本属于同一级别,相比而言带来的识别性能的提升是显著的。这说明设计的基于门控循环单元的轻量级时序序列数据分类模型,通过将遗忘门和输入门合并为一个单一的更新门,极大地简化减少了模型的参数数量,使得设计方法相对更容易训练,并最终实现对于呼吸暂停的准确检测,提出的模型能更好地适应呼吸暂停任务的诸多特点,如实时检测的迫切性以及数据集规模较小的特点。本研究选用GRU模型作为主要的分类工具,进一步对呼吸暂停严重程度进行分类,将参与者的呼吸暂停严重程度按照正常、轻度、中度和重度呼吸暂停来进行分类。经对比,提出的GRU模型在呼吸暂停严重等级上的分类准确率高达90.47%,这清晰的说明了提出模型能较好地适配呼吸暂停检测任务。

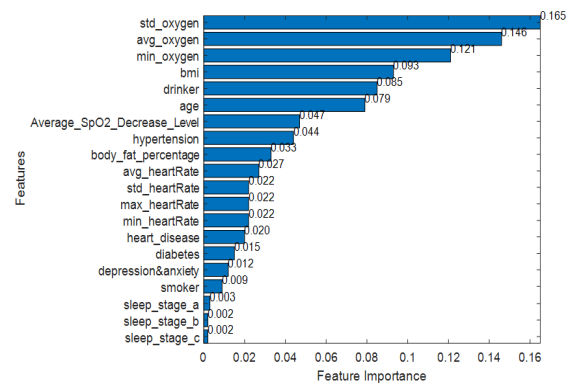


图6 特征重要性比例

Fig. 6 Proportion of feature Importance

为了深入探究特征对呼吸暂停分类结果的影响,本研究进行了特征重要性实验。针对模型性能表现最好的GRU模型,本文使用IG方法,通过计算GRU模型从基线输入到实际输入之间的路径上的梯度积分,来获取每个特征对分类结果的贡献度。通过这种方法,本文获取每个特征的重要性系数,如图6所示,图中的数值表示的是归一化之后特征重要性的相对度量,数值代表该特征对分类结果的影响在所有特征中所占比例,这些英文代表了不同的具有物理意义的特征,例如 std_oxygen 代表了血氧量的标准差, $avg_heartRate$ 代表了平均心率,所有特征重要性的总和为1(不同特征对应文章中2.1节特征工程模块)。结果表明,血氧饱和度相关数据对分类具有决定性作

用,其中血氧标准差的特征重要性得分最高,达到0.165,其次是平均血氧和血氧最小值,分别为0.146和0.121,这表明血氧饱和度水平的波动对预测结果有显著影响。当患者在发生呼吸暂停时,血氧饱和度会出现显著下降,而当呼吸暂停结束时,血氧会逐步回升到平均水平。因此,呼吸暂停发生的越频繁,血氧饱和度数据的波动则越大,这会导致血氧标准差的显著升高。另一个重要特征平均血氧体现的是试验者的血氧水平,当发生呼吸暂停,肺部供氧能力会出现下降,从而导致平均血氧水平逐步下降。实验过程中发现,健康试验者的血氧饱和度水平普遍在95%以上,而睡眠呼吸暂停患者的平均血氧水平则常常低于90%,因此该特征可以用来直观判定睡眠呼吸暂停的发生。

此外研究发现,一些病人的个人健康数据也对分类结果起到了至关重要的影响。例如身体质量指数BMI、体脂率和年龄指标也会显著影响呼吸暂停的检测结果,这些指标对模型的预测结果起到一定的贡献。BMI往往意味着更高的体脂率,从医疗角度来说,脂肪沉积会在上气道周围形成额外的压力,增加气道塌陷的风险。腹部积累的脂肪也会增加膈肌的压力,影响呼吸深度和效率。另一方面,年龄因素也与呼吸暂停显著相关,这是因为随着年龄的增长,肌肉的张力和气道结构可能会发生变化,这也可能影响气道的开放程度。这意味着随着年龄的增长,上气道肌肉会变得更加松弛,在夜间气道塌陷的可能性显著提升,从而增加了睡眠呼吸暂停的风险。因此,在评估和治疗睡眠呼吸暂停时,应该将病人减重、改善生活方式和监控健康状况作为管理睡眠呼吸暂停的一部分。

进一步地,本文采用卡方检验方法探讨类别变量与分类结果之间的相关性。所有类别变量的卡方检验 p 值均小于0.05,表明这些变量与分类结果有显著相关性。因此本文拒绝原假设,即认为这些类别变量都与分类结果相关,实验结果如图7所示。

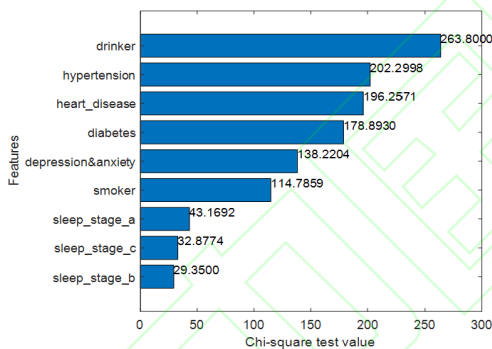


图7 类别变量的卡方检验实验结果

Fig. 7 Results of chi square test for categorical variables

卡方值越大,表示该变量与分类结果的关联性越强。可以看到关联程度比较强的四个变量是饮酒、高血压、心脏疾病和糖尿病,而三种睡眠阶段与分类结果关联程度较弱。这可以理解为饮酒、高血压以及心脏病患者相较于正常人,患有睡眠呼吸暂停的疾病的可能性更高。从医学角度来说,饮酒会导致上呼吸道肌肉放松,增加上呼吸道塌陷的风险,从而导致阻塞性睡眠呼吸暂停。此外,酒精还会影响睡眠架构和睡眠质量,增加睡眠中的呼吸异常。睡眠呼吸暂停所导致的低氧和睡眠中断会促进夜间血压升高和血压变异性增加,这最终会增加日间高血压的发生概率。心脏病可以通过多种机制增加睡眠呼吸暂停的风险。例如,心力衰竭患者可能会出现心脏功能减退和体液积聚,这些体液可以在夜间重新分布至颈部组织,进而增加上呼吸道的阻力,并迫使患者产生呼吸暂停。

最后,本文通过ANOVA和多重比较方法,分析了连续变量特征在不同分类结果间的显著性差异。实验结果显示,所有连续变量在ANOVA中的 p 值均小于0.05,表明存在显著差异。因此,本文继续使用多重比较方法来进一步探究具体哪些组别之间存在显著差异。多重比较实验结果显示,平均血氧降低水平、BMI、平均血氧、最小血氧和血氧标准差在所有的标签类别之间都有明显的差异,而年龄、体脂率、血氧最大值、心率最小值、心率标准差在部分组别之间存在显著

差异。本文通过实验结果计算标签0与标签4所代表的特征均值之间的差值,如图8所示。通过综合分析实验结果,本研究发现,血氧饱和度数据、年龄、BMI等连续变量特征,以及饮酒、高血压等类别变量特征与睡眠呼吸暂停严重程度分类结果具有较强的相关性。基于这一发现,本文进一步选取血氧饱和度数据、年龄、BMI、饮酒和高血压这几个关键特征,在之前实验的机器学习模型里综合性表现最好的RF模型上进行了精简特征集的分类实验。二次验证实验结果表明,仅依赖于这些有限的关键特征,本研究在RF模型上依然实现了80.12%的高分类准确率、78.23%的F1分数和77.65%的召回率。这一结果进一步证明了选定特征在睡眠呼吸暂停严重程度分类任务中的重要性。

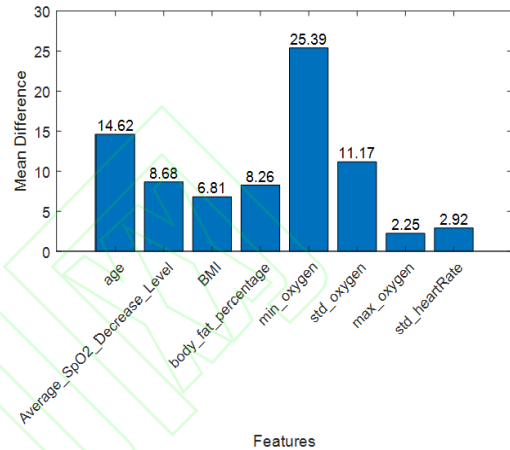


图8 连续变量的多重比较结果

Fig. 8 Multiple comparison results of continuous variables

4 结语

本文利用商用运动手环采集的数据,构建了一个多模态数据集,对睡眠呼吸暂停情况进行了检测和分类,并评估了在自然睡眠环境中使用非专业设备进行睡眠呼吸暂停监测的可行性。针对自建的手环数据集,本文基于GRU模型搭建了一个轻量级的睡眠呼吸暂停检测框架,对睡眠呼吸暂停严重程度进行了有效的分类。此外本文利用模型系数、IG方法、卡方检验、ANOVA方法和多重比较方法深入探讨了影响分类结果的关键特征变量。实验结果表明,血氧饱和度、BMI、年龄以及生活习惯如饮酒等是决定睡眠呼吸暂停分类的重要因素。

此外,本研究发现即便仅依赖于这些关键特征,也能在RF模型上实现较高的分类准确率,这一发现突显了所选定特征在睡眠呼吸暂停严重程度分类中的重要性,并为低成本和长期的睡眠障碍监测提供了新的技术途径。在未来的研究中,期望通过改进方法和增加数据集的多样性,进一步提高分类的准确性和实用性。本研究的成果不仅为睡眠呼吸暂停的自动化监测提供了新的视角,评估了智能手环在监测和诊断呼吸暂停中的潜在应用价值,也为使用机器学习方法处理医疗健康数据的研究提供了重要的参考价值。本文采集的数据集也会随着论文同步公开,以为睡眠呼吸暂停相关研究提供有效的数据支撑。在未来工作中拟进行实时的呼吸暂停检测并对未来短时间可能存在的呼吸暂停进行预测,实现对于呼吸暂停行为的预警,进一步通过的调节枕头和床体角度实现对于呼吸暂停的干预,减弱呼吸暂停给人体所带来的负面伤害。

参考文献 (References)

- [1] 吕建行, 李玉榕, 陈建国, 高宁. 两步式自适应阈值法滤除心电图运动中运动伪迹[J]. 电子学报, 2024. (LV Jian-hang, LI Yu-rong, CHEN Jian-guo, GAO Ning. ECG Motion Artifact Filtering Based on Two-Stage Adaptive Threshold Rules [J]. Acta Electronica Sinica, 2024)
- [2] Tyagi P K, Agrawal D. Automatic detection of sleep apnea from a

- single-lead ECG signal based on spiking neural network model[J]. *Computers in Biology and Medicine*, 2024, 179: 108877.
- [3] 赵津, 宋文爱, 邵隽, 杨吉江, 王青, 李晓丹, 雷毅, 邱悦. 儿童阻塞性睡眠呼吸暂停计算机人脸辅助诊断综述[J]. *计算机应用*, 2021, 41(11): 3394-3401. (Jin ZHAO, Wen'ai SONG, Jun TAI, Jijiang YANG, Qing WANG, Xiaodan LI, Yi LEI, Yue QIU. Review of computer-aided face diagnosis for obstructive sleep apnea in children[J]. *Journal of Computer Applications*, 2021, 41(11): 3394-3401.)
- [4] Lin Y, Wu Y, Lin Q, et al. Objective sleep duration and all-cause mortality among people with obstructive sleep apnea [J]. *JAMA Network Open*, 2023, 6(12): e2346085-e2346085.
- [5] Gupta K, Bajaj V, Ansari I A. OSACN-Net: automated classification of sleep apnea using deep learning model and smoothed Gabor spectrograms of ECG signal[J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 71: 1-9.
- [6] Azimi H, Bouchard M, Goubran R, et al. Unobtrusive screening of central sleep apnea from pressure sensors measurements: a patient-specific longitudinal study [J]. *IEEE Transactions on Instrumentation and Measurement*, 2020, 69(6): 3282-3296.
- [7] Thornton C S, Tsai W H, Santana M J, et al. Effects of wait times on treatment adherence and clinical outcomes in patients with severe sleep-disordered breathing: a secondary analysis of a noninferiority randomized clinical trial[J]. *JAMA Network Open*, 2020, 3(4): e203088-e203088.
- [8] Bahrami M, Forouzanfar M. Sleep apnea detection from single-lead ECG: A comprehensive analysis of machine learning and deep learning algorithms[J]. *IEEE Transactions on Instrumentation and Measurement*, 2022, 71: 1-11.
- [9] Feng K, Qin H, Wu S, et al. A sleep apnea detection method based on unsupervised feature learning and single-lead electrocardiogram [J]. *IEEE Transactions on Instrumentation and Measurement*, 2020, 70: 1-12.
- [10] Lee P, Chandel N S, Simon M C. Cellular adaptation to hypoxia through hypoxia inducible factors and beyond[J]. *Nature reviews Molecular cell biology*, 2020, 21(5): 268-283.
- [11] Wen D, Zhang X, Lei J. Consumers' perceived attitudes to wearable devices in health monitoring in China; A survey study [J]. *Computer methods and programs in biomedicine*, 2017, 140: 131-137.
- [12] Álvarez D, Cerezo-Hernández A, Crespo A, et al. A machine learning-based test for adult sleep apnoea screening at home using oximetry and airflow[J]. *Scientific reports*, 2020, 10(1): 5332.
- [13] Yook S, Kim D, Gupte C, et al. Deep learning of sleep apnea-hypopnea events for accurate classification of obstructive sleep apnea and determination of clinical severity[J]. *Sleep Medicine*, 2024, 114: 211-219.
- [14] Hu S, Liu J, Yang C, et al. Semi-supervised learning for low-cost personalized obstructive sleep apnea detection using unsupervised deep learning and single-lead electrocardiogram[J]. *IEEE Journal of Biomedical and Health Informatics*, 2023.
- [15] Yücelbaş Ş, Yücelbaş C, Tezel G, et al. Identification of full-night sleep parameters using morphological features of ECG signals: A practical alternative to EEG and EOG signals [J]. *Biomedical Signal Processing and Control*, 2024, 88: 105633.
- [16] Vaquerizo-Villar F, Gutiérrez-Tobal G C, Calvo E, et al. An explainable deep-learning model to stage sleep states in children and propose novel EEG-related patterns in sleep apnea [J]. *Computers in Biology and Medicine*, 2023, 165: 107419.
- [17] 宋钰丹, 王晶, 王雪徽, 等. 基于自适应多任务学习的睡眠生理时序分类方法[J]. *计算机应用*, 2024, 44(2): 654-662. (Yudan SONG, Jing WANG, Xuehui WANG, Zhaoyang MA, Youfang LIN. Sleep physiological time series classification method based on adaptive multi-task learning [J]. *Journal of Computer Applications*, 2024, 44(2): 654-662.)
- [18] Furlotti G. Obstructive sleep apnea detection: exploring neural activity variations through statistical analysis and machine learning [D]. *Politecnico di Torino*, 2023.
- [19] ran T D L, Nguyen T H T, Lin S Y, et al. Late Breaking Abstract- Detecting sleep apnea in advance using machine learning and ECG-deprived spectrogram during CPAP titration[J]. 2023.
- [20] Huang J, Liu B, Miao C, et al. PhyFinAtt: An Undetectable Attack Framework Against PHY Layer Fingerprint-based WiFi Authentication [J]. *IEEE Transactions on Mobile Computing*, 2023.
- [21] Abd-Alrazaq A, Aslam H, AlSaad R, et al. Detection of Sleep Apnea Using Wearable AI: Systematic Review and Meta-Analysis [J]. *Journal of Medical Internet Research*, 2024, 26: e58187.
- [22] Pillar G., Berall M., Berry R. B., Etzioni T., Henkin Y., Hwang D., Marai I., Shehadeh F., Manthena P., Rama A. N., Spiegel R., Penzel T., & Tauman R. (2022). Detection of Common Arrhythmias by the Watch-PAT: Expression of Electrical Arrhythmias by Pulse Recording. *Nature and Science of Sleep*, 14, 751 - 763.
- [23] ZHANGZ, SOWHOM, OTVOST, et al. A comparison of automated and manual sleep staging and respiratory event recognition in a portable sleep diagnostic device with in-lab sleep study [J]. *Journal of Clinical Sleep Medicine*, 2020, 16(4): 563-573.
- [24] BROWNJ, YAZDI FT, JODARI-KARIMI M, et al. Obstructive sleep apnea and hypertension: Updates to a critical relationship [J]. *Current Hypertension Reports*, 2022, 24: 173- 184.
- [25] CHEN Y, WANG W, GUO Y, et al. A single-center validation of the accuracy of a photoplethysmography-based smartwatch for screening obstructive sleep apnea [J]. *Nature and Science of Sleep*, 2021: 1533-1544.
- [26] Huang J, Liu B, Miao C, et al. PhaseAnti: An anti-interference WiFi-based activity recognition system using interference-independent phase component [J]. *IEEE Transactions on Mobile Computing*, 2021, 22(5): 2938-2954.
- [27] Penzel T, Moody G B, Mark R G, et al. The apnea-ECG database [C]//*Computers in Cardiology 2000*. Vol. 27 (Cat. 00CH3716 3). IEEE, 2000: 255-258.
- [28] O'reilly C, Gosselin N, Carrier J, et al. Montreal Archive of Sleep Studies: an open-access resource for instrument benchmarking and exploratory research[J]. *Journal of sleep research*, 2014, 23(6): 628-635.