

·运动人体科学·

## 幼儿身体活动强度识别：基于14种机器学习技术

焦坤<sup>1</sup>, 赵广高<sup>1</sup>, 付近梅<sup>2</sup>, 苏利强<sup>3</sup>, 孙顺利<sup>2</sup>, 何梓豪<sup>1</sup>, 马欣欣<sup>1</sup>,  
覃雅睿<sup>1</sup>, 刘晓虎<sup>1</sup>, 张晓雨<sup>1</sup>, 董志鑫<sup>1</sup>, 王家昌<sup>1</sup>, 韦有迪<sup>1</sup>, 贺嘉一<sup>1</sup>

(1.南昌大学 体育学院, 江西 南昌 330031; 2.江西省体育局 体育科学医疗中心, 江西 南昌 330006;  
3.江西师范大学 体育学院, 江西 南昌 330022)

**摘 要:** 运用极端梯度提升(XGB)等14种机器学习(ML)技术, 分别构建包含静坐行为(SB)以及各类幼儿身体活动(PA)强度的分类模型, 探索高准确率的幼儿PA强度识别办法。研究设计遵循COSMIN标准, 招募50名3~6岁幼儿, 在园内佩戴ActiGraph GT3X-BT并采集不同强度PA的视频数据。借助儿童PA强度评分量表(CARS)与ELAN工具将PA划分为SB、低强度PA(LPA)、中高强度PA(MVPA)3类, 或SB、LPA、中强度PA(MPA)和高强度PA(VPA)4类。以加速度计原始数据为输入特征, 采用5s、10s、15s滑动窗口切分并进行SMOTE处理, 基于XGB、支持向量机(SVM)、随机森林(RF)、卷积神经网络(CNN)、深度学习集成网络(DLEN)等14种ML技术分别构建SB-LPA-MVPA三分类和SB-LPA-MPA-VPA四分类模型。采用5折交叉验证评估各模型的表现, 并对最优模型进行等效性测试。结果显示: 在三分类任务中, 14种模型的F1分数与总体准确率分别为80.3%~97.5%与80.6%~97.5%, XGB、SVM、RF、CNN、DLEN 5种模型的F1分数与总体准确率均高于97.0%。其中, XGB的F1分数、总体准确率以及SB(99.06%±0.31%)、LPA(95.43%±1.61%)、MVPA(97.99%±0.72%)各强度类别的准确率均位居首位。在四分类任务中, 14种模型的F1分数与总体准确率分别为71.9%~97.2%与73.2%~97.2%, XGB、SVM、RF、CNN、DLEN 5种模型的F1分数与总体准确率均高于96.0%。其中, XGB的F1分数、总体准确率及SB(98.89%±0.77%)、LPA(95.51%±1.16%)、MPA(96.00%±0.72%)、VPA(98.40%±0.73%)各强度类别的准确率均居首位。等效性测试表明, XGB的预测值与实际值之间的平均偏差为-0.009 min。研究表明, XGB、SVM、RF、CNN、DLEN 5种ML技术均能够精确识别幼儿SB-LPA-MVPA与SB-LPA-MPA-VPA强度类别, 其中XGB的表现最佳。

**关键词:** 运动生理学; 幼儿身体活动; 活动强度; 机器学习; 加速度计

中图分类号: G804.2 文献标志码: A 文章编号: 1006-7116(2026)03-0139-10

### Identification of physical activity intensity in young children: Based on 14 machine learning technologies

JIAO Kun<sup>1</sup>, ZHAO Guanggao<sup>1</sup>, FU Jinmei<sup>2</sup>, SU Liqiang<sup>3</sup>, SUN Shunli<sup>2</sup>, HE Zihao<sup>1</sup>, MA Xinxin<sup>1</sup>,  
QIN Yarui<sup>1</sup>, LIU Xiaohu<sup>1</sup>, ZHANG Xiaoyu<sup>1</sup>, DONG Zhixin<sup>1</sup>, WANG Jiachang<sup>1</sup>, WEI Youdi<sup>1</sup>, HE Jiayi<sup>1</sup>

(1.School of Physical Education, Nanchang University, Nanchang 330031, China; 2.Sports Science and Medical  
Center, Jiangxi Provincial Sports Bureau, Nanchang 330006, China; 3.School of Physical Education,  
Jiangxi Normal University, Nanchang 330022, China)

**Abstract:** By using 14 machine learning (ML) techniques, from extreme gradient enhancement (XGB) to other techniques, classification models were respectively developed to cover sedentary behavior (SB) in young children and various physical activity (PA) intensity, which has designed to explore high-precision identification methods for

收稿日期: 2025-09-16

基金项目: 国家社会科学基金一般项目(25BTY097)。

作者简介: 焦坤(2000-), 男, 硕士研究生, 研究方向: 身体活动与体质健康。E-mail: 1574068774@qq.com 通信作者: 赵广高

PA intensity for this population. The study adhered to the COSMIN guidelines, and fifty children aged 3–6 years were recruited and wore ActiGraph GT3X-BT accelerometers in their preschool setting, while their PA was video-recorded. Using the Children's Activity Rating Scale (CARS) and the ELAN tool, PA was categorized into either three groups (SB, light-intensity PA [LPA], and moderate-to-vigorous PA [MVPA]) or four groups (SB, LPA, moderate PA [MPA], and vigorous PA [VPA]). Using raw accelerometer data as input features, 5 s, 10 s, and 15 s sliding windows were applied for data segmentation, followed by Synthetic Minority Over-sampling Technique (SMOTE) processing. 14 ML techniques, including XGB, support vector machine (SVM), random forest (RF), convolutional neural network (CNN), and deep learning ensemble network (DLEN), were used to construct three-class (SB-LPA-MVPA) and four-class (SB-LPA-MPA-VPA) classification models. Model performance was evaluated using 5-fold cross-validation, and an equivalence test was conducted on the optimal model. Results show that for three-class models respectively, the *F1*-scores of fourteen models change from 80.3% to 97.5%, and their total accuracy changes from 80.6% to 97.5%. Five models including XGB, SVM, RF, CNN, DLEN, have both *F1*-scores and total accuracy over 97.0%. Among them, XGB gets the first position in *F1*-score, total accuracy, and accuracy for SB, LPA, MVPA intensity classes, which are (99.06% ± 0.31%), (95.43% ± 1.61%) and (97.99% ± 0.72%) in sequence. For four-class models, *F1*-scores are in the range from 71.9% to 97.2%, and total accuracy is in the range from 73.2% to 97.2%. Therefore, the same five models, XGB, SVM, RF, CNN, DLEN, have both *F1*-scores and total accuracy over 96.0%. Hence, XGB again ranks first in *F1*-score, total accuracy, and accuracy for SB, LPA, MPA, VPA intensity classes, which are (98.89% ± 0.77%), (95.51% ± 1.16%), (96.00% ± 0.72%), (98.40% ± 0.73%), and the equivalence test results show that the average bias between XGB's predictive values and real values is -0.009 min. The conclusion holds that five machine learning techniques including XGB, SVM, RF, CNN, DLEN could carry out effective identification of SB-LPA-MVPA and SB-LPA-MPA-VPA intensity categories in young children, and among them, the XGB showing the highest performance level.

**Keywords:** sports physiology; physical activity of young children; activity intensity; machine learning; accelerometer

幼儿阶段规律的身体活动(Physical Activity, PA)对其身心健康具有重要意义<sup>[1-2]</sup>, 而 PA 的强度特征是影响其健康效应的关键因素<sup>[3]</sup>。在世界各国出台的幼儿 PA 指南中, 针对 PA 强度及时间制定了具体的推荐量标准, 如每天中高强度身体活动(MVPA)与低强度身体活动(LPA)的总时间不低于 180 min, 其中 MVPA 不低于 60 min 等<sup>[4-5]</sup>。鉴于此, 如何针对幼儿群体建立科学有效的 PA 强度识别方法, 是贯彻实施幼儿 PA 指南和保障幼儿身心健康发展的前提和基础。当前, 基于加速度计和强度切点的客观测量是体育学界常用的幼儿 PA 强度识别方法<sup>[6-8]</sup>。该方法通过对加速度计数据进行积分运算, 获取单位周期内的活动计数(count 值), 借助 Butte 等<sup>[9]</sup>、Pate 等<sup>[10]</sup>设定的强度切点(阈值), 将幼儿 PA 强度划分为高强度身体活动(VPA)、中强度身体活动(MPA)与 LPA 等。然而, 该方法会导致加速度计原始信号中大量时域、频域特征信息丢失<sup>[11]</sup>, 造成高达 35% 左右的误差, 且测量结果会因参照切点的不同产生较大差异<sup>[7]</sup>。

现今机器学习技术的出现, 为幼儿 PA 强度的精准识别提供新的方向。Ahmadi 等<sup>[12-14]</sup>使用支持向量机

(Support Vector Machine, SVM)和随机森林(Random Forest, RF), 对幼儿 PA 进行分类和性能比较。Hagenbuchner 等<sup>[15]</sup>运用深度学习集成网络(Deep Learning Ensemble Network, DLEN)和人工神经网络(Artificial Neural Network, ANN)开展对幼儿 PA 强度的分类研究。结果显示, 各模型中 LPA 的识别准确率为 68.7%~91.0%, MVPA 为 69.0%~84.0%。由此可见, 对照成人及其他群体各类模型的研究结果<sup>[16-17]</sup>, 现有幼儿 PA 强度的识别精度仍存在较大的提升空间。当前, 已有多达 14 种机器学习技术用于各类人群<sup>[12, 16-21]</sup>的 PA 强度识别中。除上述外, 还包含极端梯度提升(Extreme Gradient Boosting, XGB)、卷积神经网络(Convolutional Neural Network, CNN)、双向长短期记忆网络(Bidirectional Long Short-Term Memory Network, BLSTM)、长短期记忆网络(Long Short-Term Memory Network, LSTM)、K 近邻算法(K-Nearest Neighbors Algorithm, KNN)、决策树(Decision Tree, DT)、逻辑回归(Logistic Regression, LR)、高斯朴素贝叶斯(Gaussian Naive Bayes, GNB)、高斯混合模型(Gaussian Mixture Model, GMM)、隐马尔可夫模型(Hidden Markov Model, HMM)10 种, 就机器学习技术

的原理和PA强度的数据特征而言以上模型各有特点。

例如, XGB可通过集成决策树自动学习PA强度的多维度特征(如加速度计的均值、峰值、频率等)<sup>[19]</sup>, CNN可通过卷积核滑动提取加速度计时间序列中包含的局部周期性模式<sup>[17]</sup>, BLSTM可通过正向和反向两个LSTM层来减少识别加速度计时序特征时单一方向依赖导致的误判<sup>[16]</sup>, 而GMM可通过拟合时域、频域特征的多峰分布自动划分PA强度的概率边界等<sup>[21]</sup>。人群年龄不同、健康状态不同, 日常体力活动PA的类型分布存在差异, 加速度计采集原始信号的时域特征、频域特征也可能存在差异。在建立PA强度识别模型阶段, 目标人群年龄健康条件变动, 不同机器学习技术输出识别结果的准确程度也会发生变化。如老年人日常PA规律性偏高, 采集数据内部噪声含量低, 更适配依托统计特征建立的传统集成模型, 这类模型包括XGB、RF等<sup>[22]</sup>; 青少年日常活动频率高、动作幅度大, 捕捉这类活动的时序动态变化, 更适合深度学习模型, 常见模型包括CNN、LSTM<sup>[23]</sup>。对比健康人群, 注意缺陷多动障碍者的日常活动行为, 多呈现高频率、低持续性的特点, 发出信号更为离散、波动特征明显, 肥胖人群的活动幅度强度活动模式, 会出现异于健康人群的变化, 这些差异都会影响模型精准度和适应性<sup>[24]</sup>。生命早期阶段的健康幼儿, 活动行为天然带有碎片化短时效、高变异性特征, 加速度信号呈现非稳定性和个体差异<sup>[12]</sup>, 不同机器学习技术拥有不同算法逻辑和网络架构, 对PA强度的识别效果可能存在差异, XGB可处理非线性关系, 拥有较强泛化能力, 适配幼儿活动复杂多变的模式<sup>[25]</sup>。CNN可提取时间序列中的局部特征, 捕捉PA在时间维度上的运行模式<sup>[20]</sup>, LSTM和BLSTM可对时间序列的长短期依赖完成建模, 捕捉高频间歇性的行为特征<sup>[27]</sup>。

现有关于幼儿PA强度识别的研究, 仅覆盖RF、SVM、ANN、DLEN 4种机器学习技术<sup>[12-15, 28]</sup>, 可见系统探索并筛选幼儿PA强度识别适配度更高的模型, 已成为当下学界需要推进的研究方向。结合现有研究内容<sup>[12-15, 28]</sup>, 本研究选取上述14种机器学习技术, 建立SB和不同强度PA分类模型, 引入5s、10s、15s三种不同非重叠窗口调整识别结果, 对14种机器学习模型的性能逐一评估对比, 探索可稳定输出高准确率结果的幼儿PA强度识别模型, 为幼儿PA评价体系的完善和发展提供理论基础与技术支撑。

## 1 研究对象与方法

研究参照COSMIN指南中关于方案设计及效度量表标准化的推荐意见实施<sup>[29]</sup>。

### 1.1 研究对象

本研究对象为幼儿PA强度识别模型构建, 以3~6岁幼儿为调查对象。采用G\*power软件计算样本量, 根据Ahmadi团队<sup>[12-14]</sup>以及Hagenbuchner等<sup>[15]</sup>相关研究, 效应量为0.57~1.78, 计算函数为F-test, 设定效应量Effect Size=0.80, 显著性水平 $\alpha=0.05$ , 检验效能Power=0.90, 经测算样本量为28人。鉴于实验过程中不可避免的数据缺失问题, 实际样本量在此基础上扩充40%, 同时保留10名独立样本进行模型的等效性测试, 由此确定模型构建样本量为不少于50人。除统计功效分析外, 模型的复杂度以及泛化能力同样是关键的考量因素, 本研究选用VC维度来计算模型所需要的最小数据量, 以便能够达到最佳性能, 其公式如下:

$$N=F((VC-\ln\delta)/\varepsilon)$$

其中, 置信度 $\delta$ 设为0.05, 泛化误差 $\varepsilon$ 设为0.10, 即要求模型泛化误差不超过10%; VC维度则依据模型特征维度确定, 计算方法为特征选取数量加1。本研究共选取6个特征, 经计算发现在各类模型中一维卷积神经网络模型所需数据量最大, VC维度约为197, 数据量约为2000个, 即每种强度类别的样本数据需达到500个以上。

本研究依托江西省学龄前儿童PA测定及体质健康评价活动, 在南昌市某幼儿园开展受试者招募工作。共纳入3~6岁幼儿52名, 剔除未完成全部测试流程的样本后, 最终确定50名有效受试者(男28名, 女22名)。其中, 选取40名受试者作为训练集, 剩余10名作为独立样本进行模型的等效性测试, 样本量满足研究设计要求。当采用15s滑动窗口进行数据处理时, 训练集共获得有效数据4146个, 高于2000个的最低数据量标准。其中, 每种强度类别的样本数据为577~1626个, 均高于500个的最低样本数据要求。在研究实施前, 本方案已通过南昌大学第二附属医院医学研究伦理委员会审查。

调查对象纳入标准: 3~6岁的学龄前儿童, 身体健康, 无任何发育或运动功能障碍。所有参与儿童的父母或法定监护人需签署知情同意书, 详细说明研究目的、性质、参与的自愿性、研究流程以及隐私保护措施等。排除标准: 若儿童存在超重、肥胖问题或有生长发育障碍, 则不纳入本研究。此外, 未能完成身体形态测量项目的儿童也将被排除。

### 1.2 研究方法

1) PA强度评分量表。

儿童PA强度评分量表(Children's Activity Rating Scale, CARS)由Puhl等<sup>[30]</sup>提出, 主要采用直接观测法对2~12岁儿童的PA强度进行评估。量表划分为5个

等级, 1 级对应完全静止状态, 2 级为静坐过程中伴随四肢或躯干的轻微运动, 3~5 级对应带有位移特征的活动。每个等级都搭配明确操作定义, 列有典型活动范例, 保障评估过程准确, 也能提高评估效率; 量表可用心率和最大摄氧量校准, 帮助研究者获取更贴合实际的运动强度与能量消耗数据<sup>[30-31]</sup>。CARS 因其科学

性与实用性, 已被广泛应用于幼儿 PA 强度的评估研究中<sup>[14, 31-32]</sup>。当前学界常采用 SB-LPA-MVPA 三分类和 SB-LPA-MPA-VPA 四分类这两种 PA 强度分类方法, 1 级对应 SB, 2~3 级对应 LPA, 4~5 级对应 MVPA, 其中 4 级对应 MPA, 5 级对应 VPA<sup>[32-34]</sup>(见表 1)。

表1 幼儿 PA 水平编码规则

PA 强度分类		CARS 等级	等级描述	观察指标
三分类	四分类			
SB	SB	1 级	静止(无大肢体动作)	睡、躺、站、坐、蹲坐等
		2 级	无身体位移的四肢运动或躯干运动	站立、坐着、蹲着和跪着时伴有肢体或躯干运动如摆动、摇摆、扭转、踢腿、模仿歌曲等
LPA	LPA	3 级	缓慢简单的位移(容易)	双脚以缓慢、简单的步伐行走; 缓慢地在平面上爬行; 骑自行车
		4 级	中等速度的位移(适度)	以适中速度骑自行车; 连续跳跃、跳绳、蹦跳; 四肢协调快速爬行; 在攀爬架上爬行, 手脚并用爬上斜坡; 慢跑或者正常速度跑步; 快步走
MVPA	VPA	5 级	快速位移(困难)	全速奔跑; 全力冲刺/陡坡骑行; 负重或极限速度移动, 四肢高频率发力的爬行; 最大垂直跳、跳远(如摸高、立定跳远)

## 2) 数据采集。

(1) 视频信息采集。采用 Drift Ghost XL 便携摄像机 (Foream Network Technology Ltd., China) 在幼儿园的不同区域同步录像, 包括教室、走廊、操场等场所。室内场所安装 1 台, 固定在屋顶和墙角的交界处能拍到整个房间; 走廊因为较长就在两端高处各装 1 台; 操场上也是挑视野好的高点来拍全景, 共计 13 台。考虑到隐私问题无法在家庭环境拍摄, 在园里模拟家庭环境代替。连续录像 5 天, 每天分 8:00~12:00 和 14:00~16:30, 共计 2 轮。受试儿童佩戴加速度计于右侧髂嵴上方, 同时胸前和背后贴上号码牌以便识别。

(2) 加速度计数据采集。在为期 5 天的数据采集期内, 受试者在园内的 2 个特定时段(8:30~12:30, 14:30~16:30)佩戴 ActiGraph GT3X-BT 加速度计于右侧髂嵴上部, 设备时间与相机记录的时间保持同步。该设备可测量 3 个正交轴上的加速度, 动态范围为  $\pm 8g$ , 采样频率为 30~100 Hz。本研究采样频率设置为 60 Hz, 采用 ActiLife (Version 6.13.4) 对数据进行下载。

(3) 数据采集流程。50 名受试者被随机划分为 10 组(每组 5 人), 在每天 8:30~12:30 或 14:30~16:30 时段佩戴加速度计, 进入教室、走廊、操场等场所开展相关活动并记录数据, 连续 5 天每天 2 组。单种 PA 类型采集时长大于 5 min 且单次活动的持续时长大于 5 s 以上, 最终获得的总样本时长为 1 388.50 min, 其中, SB 458.50 min, LPA 544.60 min, MPA 194.40 min, VPA 191.00 min; 测试集共计 1 036.50 min, 其中, SB

340.00 min, LPA 406.50 min, MPA 145.75 min, VPA 144.25 min。每次活动开始之前, 研究人员都会使用统一外部时钟对笔记本电脑和相机进行时间同步校准, 以保证加速度计与视频标注时间戳的一致性。

## 3) 数据处理。

(1) 直接编码法。经过系统培训的工作人员将视频数据导入 ELAN (Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands), 结合 CARS 量表对视频中幼儿的 PA 强度进行逐段编码, 若某时间段受试儿童未出现在视频内则不对该时间段的儿童进行编码。为评估编码者之间的信度, 在视频数据中随机选取 5 个视频由两名评估员独立编码, 对于有分歧的片段研究团队共同讨论后确定最终结果, 结果表明编码的组内相关系数为 0.876(95% CI: 0.845~0.902)。

(2) 数据增强与不平衡处理。将完成直接编码后各片段的起止时间戳与加速度计原始数据进行匹配, 并采用合成少数类过采样技术 (Synthetic Minority Over-sampling Technique, SMOTE) 实现样本均衡与数据增强。对缺失值进行处理后采用 StandardScaler 对特征进行标准化, 以保证各特征处于同一尺度。SMOTE 的核心参数设置如下: random\_state=42, 采样策略采用默认的 sampling\_strategy='auto' 将少数类样本量平衡至与最大类别一致, 近邻参数 k\_neighbors=5 沿用默认设置, 相关参数已在相关文献中得到验证<sup>[35]</sup>, 可在样本真实性与多样性之间取得最佳平衡。以 15 s 窗口的数据集为例, 处理前各类别时长分布不均: SB 340.00 min,

LPA 406.50 min, MPA 145.75 min, VPA 144.25 min, 经 SMOTE 处理后各类别时长统一为 406.50 min, 实现样本均衡。

(3)数据预处理和特征提取。特征选择方面,参考前人相关研究<sup>[11, 13, 15, 36]</sup>以加速度计原始数据中的 10 th、25 th、50 th、75 th、90 th 和滞后一阶段自相关系数作为输入特征,使用配套软件 ActiLife 导出后,将三分类和四分类数据集分别划分为 5 s、10 s 和 15 s 的非重叠窗口进行分析。

(4)模型的训练和评估。本研究基于 Python 平台,利用 Scikit-learn、TensorFlow 和 XGBoost 等工具库构建 14 种机器学习模型,包括 XGB、SVM、RF、CNN、DLEN、BLSTM、ANN、LSTM、KNN、DT、LR、GNB、GMM、HMM,同时对各模型进行参数调优,进一步提高其在 PA 强度分类任务中的表现。

XGB 进行超参数优化,开展 50 次随机搜索,最优参数为  $n\_estimators=200$ ,  $max\_depth=5$ ,  $learning\_rate=0.1$ ,  $subsample=0.9$ ,  $colsample\_bytree=0.9$ ,  $reg\_alpha=0.1$ ,  $reg\_lambda=0.5$ 。SVM 采用超参数优化,利用 GridSearchCV 确定  $C=100$ ,  $kernel='rbf'$ , 并将  $gamma$  设置为“scale”或“auto”。RF 模型采用默认参数,  $n\_estimators$  设为 100,  $max\_depth=None$ ,  $in\_samples\_split$  设定为 2。CNN 经超参数优化后,采用双层卷积结构,第一层卷积核数量为 64,第二层为 128,卷积核大小均设为 2,两层卷积后分别加入 Dropout 层,丢弃率依次为 0.4 和 0.3,模型训练迭代 50 次,批量大小设置为 32。DLEN 使用 Dense(64→32)结构,Dropout 率设为 0.2,训练参数为  $epochs=50$ 、 $batch\_size=32$ 。BLSTM 由 LSTM(64,  $return\_sequences=True$ )和 LSTM(32)两层组成,并连接 Dense(64),同时加入 Dropout(0.3)。ANN 采用两层全连接结构(Dense 128→64),Dropout 率为 0.3,训练参数同 CNN。LSTM 模型采用两层结构,第一层单元数为 64,第二层降为 32,其后连接一个 64 单元的全连接层,LSTM 层后设置 Dropout 率为 0.3,迭代次数设为 50 次,批量大小为 32。KNN 超参数调优后,最优参数为  $n\_neighbors=9$ ,  $weights='distance'$ ,  $metric='euclidean'$ 。DT 采用默认参数。LR 选取  $C=1.0$ ,  $max\_iter=1\ 000$  作为最优参数。GNB 选用  $var\_smoothing=1\times 10^{-9}$ 。GMM 和 HMM 的  $n\_components=4$ ,GMM 的协方差类型选用“full”。

采用 5 折交叉验证对 14 种模型进行训练与验证,即将数据集随机划分为 5 份,每轮验证时选取 4 份作为训练集剩余 1 份作为验证集,共计完成 5 轮,该训练方法相比十折交叉验证等方法具有良好的计算效率与模型稳定性<sup>[12, 37]</sup>。采用准确度(Accuracy)、精确度

(Precision)、召回率(Recall)、F1 分数(F1-score)以及混淆矩阵(Confusion Matrix)评估模型性能,F1 分数的计算公式为:

$$F1 = 2 \times [(Precision \times Recall)/(Precision+Recall)]$$

该指标是精确度和召回率的加权调和平均值,常用来反映模型的整体性能。

为比较不同模型的分类型性能差异,本研究以最优模型为基准,对其余 13 种模型的 F1 分数进行配对  $t$  检验(paired t-test),采用 Benjamini-Hochberg 法对  $P$  值进行校正,并将校正后  $P<0.05$  作为统计学显著性判断标准,来降低多重比较导致的假阳性问题。

(5)等效性测试。为验证模型预测结果与 CARS 参考标准的一致性,本研究运用 Bland-Altman 法来对独立测试集开展分析。按照不同活动强度分别去计算幼儿累计时间的预测偏差以及相对偏差百分比,以便衡量系统误差。随后绘制 Bland-Altman 图,依据平均偏差和 95%一致性界限(Limits of Agreement)判断 2 种方法的等效程度,一致性界限越窄说明预测结果与参考标准越等效。

## 2 结果与分析

### 2.1 受试幼儿基本信息

50 名幼儿中,男童 28 例、女童 22 例,平均年龄(4.5±0.6)岁,身高(106.3±6.9)cm,体重(18.7±3.8)kg,BMI 为(16.4±2.0)kg·m<sup>-2</sup>。按性别分层后,男童的平均年龄为(4.6±0.6)岁,身高为(108.0±6.8)cm,体重为(19.2±4.2)kg,BMI 为(16.3±2.3)kg·m<sup>-2</sup>,女童的平均年龄为(4.4±0.6)岁,身高为(104.2±6.5)cm,体重为(17.9±3.0)kg,BMI 为(16.5±1.6)kg·m<sup>-2</sup>。独立样本  $t$  检验结果显示,男女幼儿在年龄、身高、体重和 BMI 方面的差异均无统计学意义。

### 2.2 模型性能

表 2 汇总 14 种模型在不同窗口大小设定下针对不同分类任务时(SB-LPA-MVPA 和 SB-LPA-MPA-VPA)的 F1 分数与准确率。结果显示,各模型在 10 s 和 15 s 窗口的 F1 分数与总体准确率均优于 5 s 窗口。

从各模型最优窗口的识别结果来看三分类任务中,14 种模型的 F1 分数与总体准确率分别为 83.3%~97.5%与 83.7%~97.5%,XGB、SVM、RF、CNN、DLEN 5 种模型的 F1 分数与总体准确率均高于 97.0%。其中,XGB 的 F1 分数和总体准确率均具首位。四分类任务中,14 种模型的 F1 分数与总体准确率分别为 73.6%~97.2%与 75.2%~97.2%,XGB、SVM、RF、CNN、DLEN 5 种模型的 F1 分数与总体准确率均高于 96.0%,且 XGB 的 F1 分数和总体准确率均具首位。

表2 模型分类结果<sup>1)</sup>

指标	F1						准确率					
	三分类			四分类			三分类			四分类		
	5 s	10 s	15 s	5 s	10 s	15 s	5 s	10 s	15 s	5 s	10 s	15 s
XGB	0.966	0.973	0.975	0.961	0.970	0.972	0.966	0.974	0.975	0.961	0.970	0.972
SVM	0.956	0.972	0.971	0.958	0.968	0.971	0.957	0.972	0.971	0.958	0.968	0.971
RF	0.962	0.971	0.971	0.954	0.961	0.963	0.962	0.971	0.971	0.954	0.962	0.963
CNN	0.957	0.969	0.97	0.955	0.962	0.964	0.958	0.968	0.97	0.955	0.962	0.963
DLEN	0.956	0.969	0.975	0.944	0.955	0.961	0.956	0.969	0.975	0.944	0.955	0.961
BLSTM	0.956	0.966	0.966	0.945	0.952	0.951	0.956	0.966	0.966	0.945	0.952	0.951
ANN	0.956	0.969	0.974	0.93	0.947	0.949	0.956	0.969	0.974	0.930	0.947	0.949
LSTM	0.954	0.961	0.968	0.927	0.94	0.932	0.953	0.961	0.968	0.927	0.940	0.931
KNN	0.926	0.946	0.951	0.934	0.944	0.962	0.927	0.947	0.952	0.935	0.944	0.962
DT	0.931	0.935	0.931	0.91	0.919	0.921	0.932	0.935	0.931	0.91	0.919	0.921
LR	0.919	0.934	0.942	0.854	0.879	0.885	0.919	0.935	0.942	0.855	0.880	0.886
GNB	0.803	0.823	0.833	0.719	0.739	0.736	0.806	0.827	0.837	0.732	0.753	0.752
GMM	0.803	0.823	0.833	0.719	0.739	0.736	0.806	0.827	0.837	0.732	0.753	0.752
HMM	0.803	0.823	0.833	0.719	0.739	0.736	0.806	0.827	0.837	0.732	0.753	0.752

1)表中数字均为均值。

表3为15 s非重叠窗口下,13种模型与XGB的配对 $t$ 检验结果。三分类任务中,XGB显著优于8种

模型( $P < 0.05$ )。在四分类任务中,XGB显著优于11种模型( $P < 0.05$ )。

表3 各模型最优结果与XGB的配对 $t$ 检验结果

任务	模型	F1	$t$	未校正 $P$ 值	校正 $P$ 值	效应量
3-15	SVM	0.971±0.006	1.200	0.295	0.348	0.701
	RF	0.971±0.003	1.880	0.134	0.193	1.168
	CNN	0.970±0.005	1.670	0.170	0.221	1.106
	DLEN	0.975±0.003	0.060	0.957	0.957	0.044
	BLSTM	0.966±0.005	4.930	0.008	0.013	1.943
	ANN	0.974±0.003	0.670	0.541	0.586	0.309
	LSTM	0.968±0.003	7.230	0.002	0.004	1.942
	KNN	0.951±0.003	13.890	0.000	0.000	6.989
	DT	0.931±0.013	6.640	0.003	0.005	4.235
	LR	0.942±0.005	17.060	0.000	0.000	6.969
	GNB	0.833±0.014	18.700	0.000	0.000	12.398
	GMM	0.833±0.014	18.700	0.000	0.000	12.398
	HMM	0.833±0.014	18.700	0.000	0.000	12.398
4-15	SVM	0.971±0.003	0.330	0.755	0.755	0.122
	RF	0.963±0.004	5.640	0.005	0.008	1.729
	CNN	0.964±0.006	2.800	0.049	0.053	1.336
	DLEN	0.961±0.007	4.230	0.013	0.017	1.609
	BLSTM	0.951±0.004	10.510	0.001	0.001	3.886
	ANN	0.949±0.004	9.290	0.001	0.001	4.227
	LSTM	0.932±0.017	5.230	0.006	0.009	2.890
	KNN	0.962±0.002	3.170	0.033	0.040	2.353
	DT	0.921±0.008	18.800	0.000	0.000	6.770
	LR	0.885±0.005	56.570	0.000	0.000	14.481
	GNB	0.736±0.012	54.500	0.000	0.000	23.310
	GMM	0.736±0.012	54.500	0.000	0.000	23.310
	HMM	0.736±0.012	54.500	0.000	0.000	23.310

图1最优模型XGB 15 s窗口下混淆矩阵。三分类时见图1左,SB的准确率为99.1%、LPA为95.4%、MVPA为98.0%。其中,0.9%的SB误判为LPA,2.9%与1.6%的LPA分别误判为MVPA与SB,2.0%的MVPA误判为LPA。四分类时见图1右,SB的准确率为

98.9%、LPA为95.5%、MPA为96.0%、VPA为98.4%。其中,1.1%的SB误判为LPA,1.4%与3.1%的LPA分别误判为SB与MPA,0.9%与3.1%的MPA分别误判为LPA与VPA,1.6%的VPA误判为MPA。

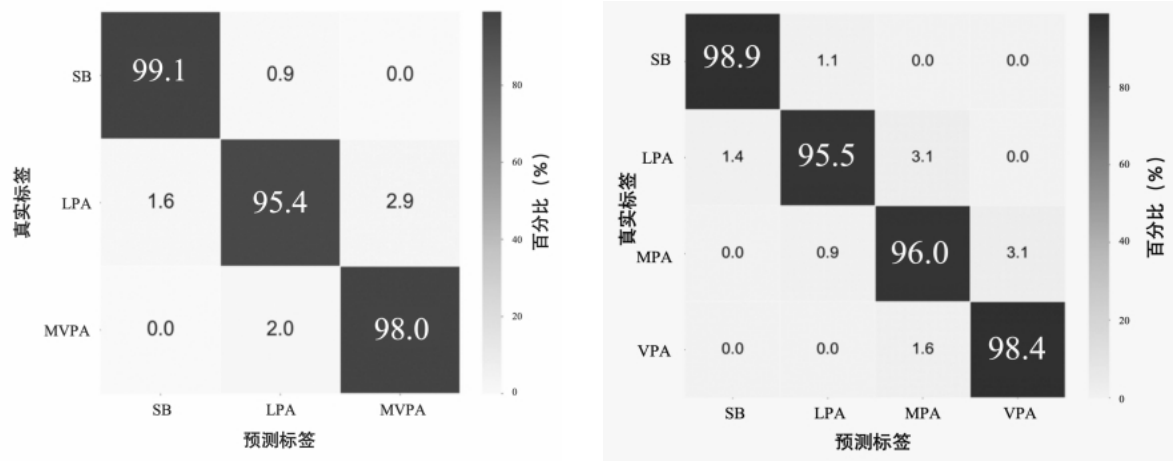


图1 XGB 混淆矩阵

### 2.3 等效性测试

在最优模型 XGB 与 CARS 的等效性测试中, 各类别的平均时间分布为, SB 的 CARS 测量值为 11.85 min, XGB 预测值为 12.06 min, 二者差异为 0.21 min, 差异百分比为 +1.80%。LPA 的 CARS 测量值为 13.81 min, XGB 预测值为 13.70 min, 差异为 -0.11 min, 差异百分比为 -0.80%。MVPA 的 CARS 测量值为 9.54 min, XGB 预测值为 9.44 min, 差异为 -0.10 min, 差异百分比为 -1.00%。Bland-Altman 分析显示, 模型预测值与 CARS 测量值的平均偏差为 -0.009 min, 95.1% 的数据点位于一致性界限内, 一致性区间宽度为 0.876 min。

## 3 讨论

本研究基于 14 种机器学习技术构建幼儿 PA 强度识别数据模型。在已有研究中, 学者们<sup>[12-15, 28]</sup>借助其中的 RF、SVM、ANN、DLEN 4 种机器学习技术开展同类研究工作。相比已有普遍采用的 SB-LPA-MVPA-步行-跑步五分类方法<sup>[12-15, 28]</sup>, 为了与加速度计传统强度切点<sup>[6, 32, 38-39]</sup>和指南标准<sup>[4]</sup>对应, 本研究分别采用 SB-LPA-MVPA 三分类和 SB-LPA-MPA-VPA 四分类任务进行模型构建。结果显示, 在 14 种模型中三分类与四分类模型的准确率分别为 80.6%~97.5% 与 73.2%~97.2%, *F1* 分数分别为 80.3%~97.5% 与 71.9%~97.2%, 而前人研究各模型的准确率为 59.1%~85.0%, *F1* 分数均未报道。其中, XGB、SVM、RF、CNN、DLEN 5 种模型对于三分类与四分类任务的准确率均分别高于 97.0% 与 96.0%, XGB 均具首位, 而前人研究的准确率则均未超过 85.0%。SB-LPA-MVPA 三分类时, XGB 等 5 种模型各 PA 类别的最优准确率分别为 SB(99.0%~99.1%)、LPA(92.7%~95.4%) 与 MVPA(90.7%~98.0%)。与既往研

究(SB 为 76.8%~94.7%, LPA 为 71.1%~91.0%, MVPA 为 79.0%~83.6%) 相比, 本研究在 SB、LPA 和 MVPA 上的准确率分别提升 4.4%~22.2%、4.4%~21.6% 和 11.7%~14.4%。配对 *t* 检验结果呈现 XGB 的性能显著优于 8 种模型( $P < 0.05$ ), 等效性分析结果说明, XGB 预测值跟 CARS 人工标识值的平均偏差是 -0.009 min, 95.1% 的数据点处在 95% 一致性区间之内, 一致性区间的宽度为 0.876 min 处于临床范围里面, 这表明 XGB 预测结果与 CARS 人工标识值之间有着高度一致性。本研究识别准确率的提高, 主要原因包括模型选择、类别划分策略、数据预处理以及样本一致性等因素。

本研究对比 14 种机器学习模型, 结果显示, XGB 的表现优于 SVM、RF、CNN、DLEN 等其余 13 种模型, XGB 的优势可从多个维度梳理, XGB 属于基于树的集成学习方法, 可以高效捕捉各类特征之间复杂的非线性关系, 借助梯度提高策略不断调整模型输出效果<sup>[40]</sup>。XGB 内部自带正则化机制, 可抑制过拟合问题出现, 适配样本量有限、特征维度偏高的 PA 数据<sup>[40, 41]</sup>。XGB 在本次研究中表现优于 SVM, 原因大概率指向 SVM 本身特性, SVM 对核函数和参数设置敏感度高, 特征维度偏高或特征之间存在复杂非线性关系时分类性能会受到限制。XGB 依托梯度提高机制, 可自动为多维特征之间的非线性关系建立模型, 不需要引入复杂核函数就能完成模型优化<sup>[42]</sup>。RF 同样属于树模型, XGB 的准确率和泛化能力通常优于 RF<sup>[41]</sup>, CNN、BLSTM、LSTM、DLEN 都属于深度学习模型。和这类模型相比, XGB 对数据预处理和特征工程的依赖更低, 对样本量的要求更少, 幼儿活动数据样本量普遍有限, 这一特性在此场景下发挥作用<sup>[41, 43]</sup>。深度学习模型可自动提取时序特征和空间特征, 训练过程复杂且对计算资源需求量大, 训练结果容易出现过拟合, XGB

在偏小的数据集中可快速完成收敛,最终获得高性能输出<sup>[41,43]</sup>。

现有研究<sup>[12-15,28]</sup>对 PA 输出类别展开划分时,多将 PA 归为 SB、LPA、MVPA、步行、跑步几类,本研究聚焦两种划分方式且均以活动强度为核心,分别为 SB-LPA-MVPA 与 SB-LPA-MPA-VPA,具体行为存在动作差异,不同个体也存在变化。本研究采用的划分方法可避开这些因素带来的复杂性,帮助提高模型泛化能力,也可提高分类准确性<sup>[44]</sup>。研究将步行跑步作为单独动作类别识别时,对应特征可能与 LPA、MVPA 的特征存在重叠,模型无法准确区分不同类别,整体分类效果会受到影响<sup>[45]</sup>,采用仅依托强度的划分方式可简化分类任务,也可提高模型实际应用中的稳定性,拓展实用价值<sup>[44-45]</sup>。

由于原始数据分布不均衡(SB 340.00 min, LPA 406.50 min, MPA 145.75 min, VPA 144.25 min),本研究使用 SMOTE 技术对少数类样本进行合成过采样以实现数据扩增和类别平衡。相较于把少数类样本直接进行复制的过采样办法,SMOTE 凭借在特征空间里对近邻样本开展插值操作来生成新的合成样本,提高少数类样本的多样性,处理模型过拟合的问题。另外这种方法可以把数据在特征空间中的局部结构予以保留,让生成的样本更契合原始数据分布,这样可以避免出现无效样本的情况。在高维特征空间中,如加速度计采集的 PA 数据,SMOTE 扩展少数类的决策边界,帮助模型更准确地区分各个类别。研究表明,SMOTE 与基于树的集成学习模型(如 XGB 和 RF)结合使用时能够显著提高少数类的召回率和 F1 分数<sup>[40]</sup>。

数据一致性方面,PA 类型差异为主导因素。如 Hagenbuchner 等<sup>[15]</sup>收集看电视、朗读、使用平板电脑、拼贴画、寻宝、清理玩具、骑自行车、跑步活动,这些任务多以结构化、短时限、可控性强的活动为主。本研究所选用的数据源于幼儿在教学以及游戏场景当中的连续活动,涉及到更多集体性以及动态性的动作,致使本研究的动作类型跟其他相关研究有着一定差异,可能会对模型在不同数据集之间的可迁移性以及泛化能力产生影响。

本研究在实践应用和科学研究方面均具有较高的拓展价值。在实际应用中,机器学习模型能够从多维时域特征当中开展分类信息的提取工作,来契合幼儿运动行为所有的短时、碎片化以及类型多样这些特性,和传统固定阈值切点法相比拥有更高的识别效率,将来可以借助“数据采集—特征提取—模型训练—实时识别”的流程,来搭建适宜于幼儿园以及家庭场景的幼儿体力活动强度智能识别系统。高精度的体力活动

强度分类结果,能够为幼儿园课程结构的优化工作、运动环境的配置工作以及个性化活动处方的调整工作,提供量化方面的支撑,同时也可以为幼儿体力活动指南的制定以及公共卫生政策的循证决策,提供数据方面的参考。针对肥胖、运动发育迟缓以及注意缺陷多动障碍等特殊群体而言,他们的活动模式、姿态控制以及能量消耗都有着较大差异。将来能够去开展针对性数据集的创建工作,设计出更具鲁棒性的识别算法,来实现对该类人群运动评估、行为监测以及康复干预的支持。在科学研究这个方面,本研究鉴于幼儿 PA 强度与健康效应之间的关系,提供更精准的方法路径。以往有关 PA 强度与体质健康、运动发展或者认知行为之间关系的研究大多运用切点法,识别精度的局限性或许会影响效应推断的可靠性。凭借机器学习技术更为精细且契合幼儿真实活动模式的强度分布,能够极大程度上提高相关性分析以及因果推断的有效性,将来研究可以把基于机器学习的强度分类结果跟纵向健康指标整合起来,从更具机制化以及发展性的角度去揭示 PA 在推动体质发展、运动技能形成以及认知行为调节等潜在作用机制,让幼儿 PA 研究朝着更精细化、数据驱动以及理论深化的方向去发展。

本研究严格遵循科学研究方法进行,但仍存在若干局限。其一,各 PA 强度数据分布不均,尤其是 MPA 和 VPA 样本较少。但要说明的是,本研究采集的数据量已超过机器学习模型构建所需的样本数,并且借助数据提高技术达成样本数量的平衡。其二,鉴于隐私保护和安全方面的限制,研究未能在家庭、社区、户外等自然环境中录制视频以获取相关活动数据。然而在幼儿园环境当中,本研究已经尽力包含家庭生活里涉及的多种强度 PA,以此来保证模型的代表性与适宜性。

## 参考文献:

- [1] World Health Organization. Guidelines on physical activity, sedentary behaviour and sleep for children under 5 years of age[M]. Geneva: World Health Organization, 2020.
- [2] 郭梁,李天庚,曾赛,等. 学龄儿童 24 h 活动时间分配和替代与操作智商的关系[J]. 体育学刊, 2025, 32(4): 144-150.
- [3] LU Z, GUO J, LIU C, et al. Reallocation of time to moderate-to-vigorous physical activity and estimated changes in physical fitness among preschoolers: A compositional data analysis[J]. BMC Public Health, 2024, 24(1): 2823.
- [4] WILLUMSEN J, BULL F. Development of WHO guidelines on physical activity, sedentary behavior, and sleep for children less than 5 years of age [J]. J Phys Act Health,

- 2020, 17(1): 96-100.
- [5] 《中国妇女发展纲要(2021—2030年)》《中国儿童发展纲要(2021—2030年)》[N]. 2021-09-28(003).
- [6] 胡学文, 赵广高, 付近梅, 等. 基于等时替代模型的幼儿身体活动、久坐与睡眠的关系[J]. 体育学刊, 2024, 31(5): 143-150.
- [7] LEPPÄNEN M H, MIGUELES J H, ABDOLLAHI A M, et al. Comparing estimates of physical activity in children across different cut-points and the associations with weight status[J]. Scand J Med Sci Sports, 2022, 32(6): 971-983.
- [8] 邱艳平, 王丽娟, 周玉兰, 等. 基于成分数据分析的24h活动与儿童基本动作技能的关系[J]. 体育学刊, 2023, 30(1): 137-144.
- [9] BUTTE N F, WONG W W, LEE J S, et al. Prediction of energy expenditure and physical activity in preschoolers[J]. Med Sci Sports Exerc, 2014, 46(6): 1216-1226.
- [10] PATE R R, ALMEIDA M J, MCIVER K L, et al. Validation and calibration of an accelerometer in preschool children[J]. Obesity (Silver Spring), 2006, 14(11): 2000-2006.
- [11] 陈庆果, 彭彪, 杨世军, 等. 基于神经网络模型的加速度计活动强度算法研究[J]. 天津体育学院学报, 2017, 32(1): 45-50+80.
- [12] AHMADI M N, TROST S G. Device-based measurement of physical activity in pre-schoolers: Comparison of machine learning and cut point methods[J]. PLoS One, 2022, 17(4): e0266970.
- [13] TROST S G, CLIFF D P, AHMADI M N, et al. Sensor-enabled activity class recognition in preschoolers: Hip versus wrist data[J]. Med Sci Sports Exerc, 2018, 50(3): 634-641.
- [14] AHMADI M N, BROOKES D, CHOWDHURY A, et al. Free-living evaluation of laboratory-based activity classifiers in preschoolers [J]. Med Sci Sports Exerc, 2020, 52(5): 1227-1234.
- [15] HAGENBUCHNER M, CLIFF D P, TROST S G, et al. Prediction of activity type in preschool children using machine learning techniques[J]. Journal of Science and Medicine in Sport, 2015, 18(4): 426-431.
- [16] NG J Y Y, ZHANG J H, HUI S S, et al. Development of a multi-wear-site, deep learning-based physical activity intensity classification algorithm using raw acceleration data[J]. PLoS One, 2024, 19(3): e0299295.
- [17] FARRAHI V, MUHAMMAD U, ROSTAMI M, et al. AccNet24: A deep learning framework for classifying 24-hour activity behaviours from wrist-worn accelerometer data under free-living environments[J]. Int J Med Inform, 2023, 172: 105004.
- [18] KHATAEIPOUR S J, ANARAKI J R, BOZORGIA, et al. Predicting lying, sitting and walking at different intensities using smartphone accelerometers at three different wear locations: Hands, pant pockets, backpack[J]. BMJ Open Sport Exerc Med, 2022, 8(2): e001242.
- [19] BAI C, WANIGATUNGA A A, SALDANA S, et al. Are machine learning models on wrist accelerometry robust against differences in physical performance among older adults?[J]. Sensors (Basel), 2022, 22(8): 1742.
- [20] O'DRISCOLL R, TURICCHI J, HOPKINS M, et al. Comparison of the validity and generalizability of machine learning algorithms for the prediction of energy expenditure: Validation study[J]. JMIR Mhealth Uhealth, 2021, 9(8): e23938.
- [21] DUTTA A, MA O, TOLEDO M, et al. Identifying free-living physical activities using lab-based models with wearable accelerometers[J]. Sensors, 2018, 18(11): 2107.
- [22] MARDINI M T, BAI C, WANIGATUNGA A A, et al. Age differences in estimating physical activity by wrist accelerometry using machine learning[J]. Sensors, 2021, 21(10): 3352.
- [23] LIANG Y T, WANG C, HSIAO C K. Data analytics in physical activity studies with accelerometers: Scoping review[J]. J Med Internet Res, 2024, 26: e59497.
- [24] PETERSEN B A, ERICKSON K I, KUROWSKI B G, et al. Emerging methods for measuring physical activity using accelerometry in children and adolescents with neuromotor disorders: A narrative review[J]. J Neuroeng Rehabil, 2024, 21(1): 31.
- [25] CHEN T, GUESTRIN C. XGBoost: A scalable tree boosting system[J]. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, 38(6): 1728.
- [26] SERGHINI O, SEMLALI H, MAALI A, et al. 1-D convolutional neural network-based models for cooperative spectrum sensing[J]. Future Internet, 2024, 16(1): 14.
- [27] SAHA S, MARTIN C, VALAFAR H. Human Activity Recognition on Time Series Accelerometer Sensor Data using LSTM Recurrent Neural Networks[M]. 2022.
- [28] AHMADI M N, PAVEY T G, TROST S G. Machine learning models for classifying physical activity in

- free-living preschool children[J]. *Sensors*, 2020, 20(16): 4196.
- [29] MOKKINK L B, TERWEE C B, PATRICK D L, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes[J]. *J Clin Epidemiol*, 2010, 63(7): 737-745.
- [30] PUHL J, GREAVES K, HOYT M, et al. Children's activity rating scale (CARS): description and calibration[J]. *Res Q Exerc Sport*, 1990, 61(1): 26-36.
- [31] VAN CAUWENBERGHE E, LABARQUE V, TROST S G, et al. Calibration and comparison of accelerometer cut points in preschool children[J]. *Int J Pediatr Obes*, 2011, 6(2-2): e582-589.
- [32] 常振亚, 王树明. 活动计数在学龄前儿童体力活动强度诊断中最佳临界值的校准、验证与运用[J]. *首都体育学院学报*, 2021, 33(1): 74-83.
- [33] JOHANSSON E, LARISCH L M, MARCUS C, et al. Calibration and validation of a wrist- and hip-worn actigraph accelerometer in 4-year-old children[J]. *PLoS One*, 2016, 11(9): e0162436.
- [34] JANSSEN X, CLIFF D P, REILLY J J, et al. Predictive validity and classification accuracy of ActiGraph energy expenditure equations and cut-points in young children[J]. *PLoS One*, 2013, 8(11): e79124.
- [35] FERNÁNDEZ A, GARCÍA S, HERRERA F, et al. SMOTE for learning from imbalanced data[J]. *Journal of Artificial Intelligence Research*, 2018, 46(2): 1082.
- [36] TROST S G, WONG W-K, PFEIFFER K A, et al. Artificial neural networks to predict activity type and energy expenditure in youth[J]. *Medicine and Science in Sports and Exercise*, 2012, 44(9): 1801-1809.
- [37] CHEN D, DU Y, LIU Y, et al. Development and validation of a smartwatch algorithm for differentiating physical activity intensity in health monitoring[J]. *Sci Rep*, 2024, 14(1): 9530.
- [38] 常振亚, 王树明. 学龄前儿童身体活动采样间隔和强度分界值的适用性研究[J]. *中国体育科技*, 2022, 58(6): 17-25.
- [39] 赵广高, 洪熊, 苏利强, 等. 高强度间歇训练与中强度持续训练对超重/肥胖成人成分干预作用的比较: Meta分析[J]. *成都体育学院学报*, 2017, 43(4): 93-99.
- [40] CHEN T, GUESTRIN C. XGBoost: A Scalable Tree Boosting System[Z]. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, California, USA; Association for Computing Machinery, 2016.
- [41] GRINSZTAJN L, OYALLON E, VAROQUAUX G. Why do tree-based models still outperform deep learning on typical tabular data?[Z]. *Proceedings of the 36th International Conference on Neural Information Processing Systems*. New Orleans, LA, USA; Curran Associates Inc. 2022.
- [42] SARKER P, TIANG J J, NAHID A A. Metaheuristic-driven feature selection for human activity recognition on KU-HAR dataset using XGBoost classifier[J]. *Sensors*, 2025, 25(17): 4189.
- [43] ALAGOZ C. Comparative analysis of XGBoost and minirocket algorithms for human activity recognition[J]. *ArXiv*, 2024, 24(2): 18296.
- [44] TROST S G, BROOKES D S K, AHMADI M N. Evaluation of wrist accelerometer cut-points for classifying physical activity intensity in youth[J]. *Front Digit Health*, 2022, 4: 884307.
- [45] RICO-GONZÁLEZ M, GÓMEZ-CARMONA C D. Machine learning applications for physical activity and behaviour in early childhood: A systematic review[J]. *Applied Sciences*, 2025, 15(11): 6296.