

精简惯性测量单元时序模型的冻结步态患者级早期识别*

赵乙璇, 高瑞, 任依飒, 肖启睿, 王玉锋, 高建强, 李莉*

(济宁医学院医学信息工程学院, 山东 日照 276800)

【摘要】目的/意义 构建面向未见患者的可穿戴惯性测量单元冻结步态早期识别模型, 评估其跨患者稳定性以及关键特征和网络模块的作用。**方法/过程** 基于 12 名患者的 56 条公开记录, 采用传感器可用性核查和方向无关小腿特征汇聚, 构建融合残差一维卷积、双向长短期记忆网络和注意力池化的精简惯性测量单元时序模型, 并采用 3 个固定随机种子进行患者级留一内部验证和消融分析。**结果/结论** 模型的准确率、平衡准确率、F1 值、AUROC 和 AUPRC 分别为 0.7910、0.7378、0.6325、0.8791 和 0.7595。冻结一步态频带特征及双向时序建模对模型性能贡献较大, 在本数据集的患者级留一验证中表现出一定的跨患者稳定性。

【关键词】 帕金森病; 冻结步态; 惯性测量单元; 早期识别; 患者级交叉验证

【中图分类号】 R74

【基金项目】 教育部产学合作协同育人项目“人工智能技术融入《信号与系统》课程教学的路径探索”(项目编号: 2511173428); 济宁医学院 2020 年教育教学研究项目“基于移动互联网络的生物医学工程专业自主深度学习模式研究”(项目编号: Y2020037); 济宁医学院教师科研扶持基金(项目编号: JYFC2019KJ014); 济宁医学院博士科研启动基金(项目编号: 2018JYQD03); 济宁医学院 2025 年大学生创新训练计划项目(项目编号: cx2025121)。

【作者简介】 赵乙璇, 学生。通信作者: 高建强, 博士, 副教授; 李莉, 博士, 讲师, 主要研究方向为模式识别与智能系统, email: liliot@163.com。

Patient-Level Early Identification of Freezing of Gait Using a Streamlined Temporal Model Based on Inertial Measurement Unit Data

Yixuan Zhao, Rui Gao, Yisa Ren, Qirui Xiao, Yufeng Wang, Jianqiang Gao, Li Li

(School of Medical Information Engineering, Jining Medical University, Rizhao 276800, China)

【Abstract】 Objective/Significance This study presents a wearable inertial measurement unit (IMU)-based framework for early freezing of gait (FOG) identification in previously unseen patients, specifically targeting cross-patient generalizability and dissecting the relative contributions of key spectral features and network modules. **Methods/Process** We interrogated a

publicly available dataset comprising 56 recordings from 12 patients. After verifying sensor availability, we performed direction-invariant feature aggregation from shank-mounted IMU signals. A compact temporal architecture was subsequently designed, integrating residual one-dimensional convolutions, bidirectional long short-term memory (LSTM) networks, and an attention-based pooling mechanism. Internal validation employed a patient-level leave-one-subject-out (LOSO) scheme, alongside systematic ablation studies, all conducted across three fixed random seeds. **Results/Conclusion** The proposed pipeline yielded an accuracy of 0.7910, a balanced accuracy of 0.7378, an F1-score of 0.6325, an area under the receiver operating characteristic curve (AUROC) of 0.8791, and an area under the precision–recall curve (AUPRC) of 0.7595. The freeze- and gait-band spectral features, together with bidirectional temporal modeling, emerged as primary drivers of performance. Notably, across the three fixed seeds during LOSO internal validation on the current dataset, the AUROC and AUPRC exhibited only limited variability, confirming reliable cross-patient stability.

[Keywords] Parkinson’s disease; freezing of gait; inertial measurement unit; early identification; patient-level cross-validation

1 引言

帕金森病患者在起步、转弯或经过狭窄通道时，可能突然出现步伐中断或难以继续前进的现象，即冻结步态（freezing of gait, FOG）。这类疾病发作会增加跌倒及日常活动受限的风险^[1]。惯性测量单元（inertial measurement unit, IMU）可记录加速度和角速度，因而被用于 FOG 的可穿戴监测^[2]。然而，现有研究在传感器位置、任务设置、标签边界和评价层级上差异明显，同一个“预测”指标可能分别指向 FOG 发生前识别、已发生 FOG 检测或患者依赖分类，不同研究结果难以直接比较^[3]。

现有方法大致分为特征工程与深度时序模型两类。特征工程方法通常从运动幅度、步态频带和冻结频带中提取指标描述 FOG 相关变化^[4]。近期基于可穿戴传感器的算法研究进一步评估了 FOG 检测与发作前预测能力，但较高的假阳性率仍是临床应用的重要限制^[5]。面向患者特异性的动态建模也显示出一定的 FOG 提前识别潜力，但其性能仍依赖患者数据和个体阈值设置^[6]。深度时序模型则利用双向时序网络和通道选择方法分析多部位运动信号，以提高 FOG 事件识别效率^[7]，并通过多通道时序网络增强跨时间依赖建模^[8]。不同网络架构

之间的比较还显示，数据划分和任务定义有时比增加网络复杂度更能影响最终指标^[9]。因此，本研究将评估重点放在患者完全隔离条件下的特征贡献与网络结构影响上，而非单纯比较更复杂的模型架构。

本研究围绕三个实际问题展开：小腿传感器部分失效时如何构建统一维度的 IMU 输入；哪些特征和时序结构在未见患者中仍能保持有效性；困难患者的性能下降是否与随机初始化或患者间分布差异有关。

基于上述问题，本研究使用公开多模态 FOG 数据。每条记录先确定有效小腿传感器，再从有效侧信号中构建方向无关的统一输入。实验在患者级留一框架下完成，分别比较特征组、残差卷积、BiLSTM 和注意力池化的影响，同时记录动态特征扩展、sEMG 融合及个体化校准的结果。患者 009 被单独用于分析低敏感度的形成原因。各折的预处理参数、早停和模型选择均不使用外层测试患者的数据。

2 资料与方法

2.1 数据来源与样本分析

数据来源于 Zhang 等发布的帕金森病冻结步态（freezing of gait, FOG）多模态公开数据集^[10]。本研究纳入 001—012 号 12 名患者的 56 条有效记录。以输入序列末端窗口的预测时刻确定类别标签：若该时刻位于某次 FOG 发作起始前 2.0 s 至发作结束之间，则标记为阳性；若位于无 FOG 事件影响的有效背景时段，则标记为阴性；FOG 结束后 1.0 s 内的恢复期作为灰区，不参与模型训练和主要性能评价。经窗口筛选和序列构建后，共获得 23636 条有效序列，其中阳性序列 11742 条，阴性序列 11894 条。患者及序列构成见表 1。

表 1 患者记录和序列构成

患者	记录时长/min	记录文件数	序列数	阳性序列数	阳性率
001	12.550	4	1422	608	42.8%
002	8.950	4	1032	10	1.0%
003	34.480	4	3996	3649	91.3%
004	10.730	5	1208	325	26.9%
005	5.100	4	572	0	0.0%
006	17.680	4	2038	800	39.3%
007	11.430	4	1272	765	60.1%

患者	记录时长/min	记录文件数	序列数	阳性序列数	阳性率
008	29.470	9	3352	1907	56.9%
009	13.200	6	1448	372	25.7%
010	26.780	4	3104	1664	53.6%
011	19.350	4	2180	1104	50.6%
012	17.300	4	2012	538	26.7%

注：通道映射来自数据集元数据；患者 005 无 FOG 事件。

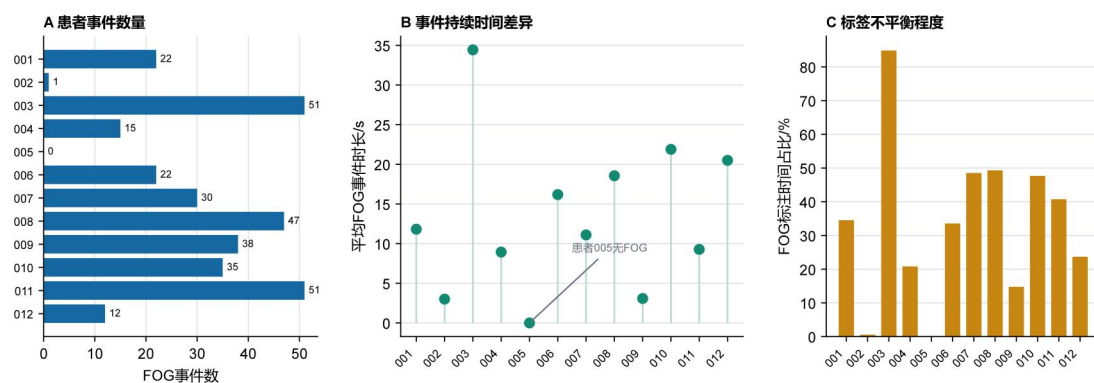


图 1 患者级 FOG 事件与标注时长分布

2.2 研究设计与总体技术路线

图 2 给出了从原始记录检查到患者级评价的处理过程。每个外层折完整留出 1 名患者作为测试对象，其余患者进一步划分为 1 名验证患者和若干训练患者。缺失值填补、异常值截断、特征删除、标准化参数、类别权重及早停轮次均由训练侧数据确定。

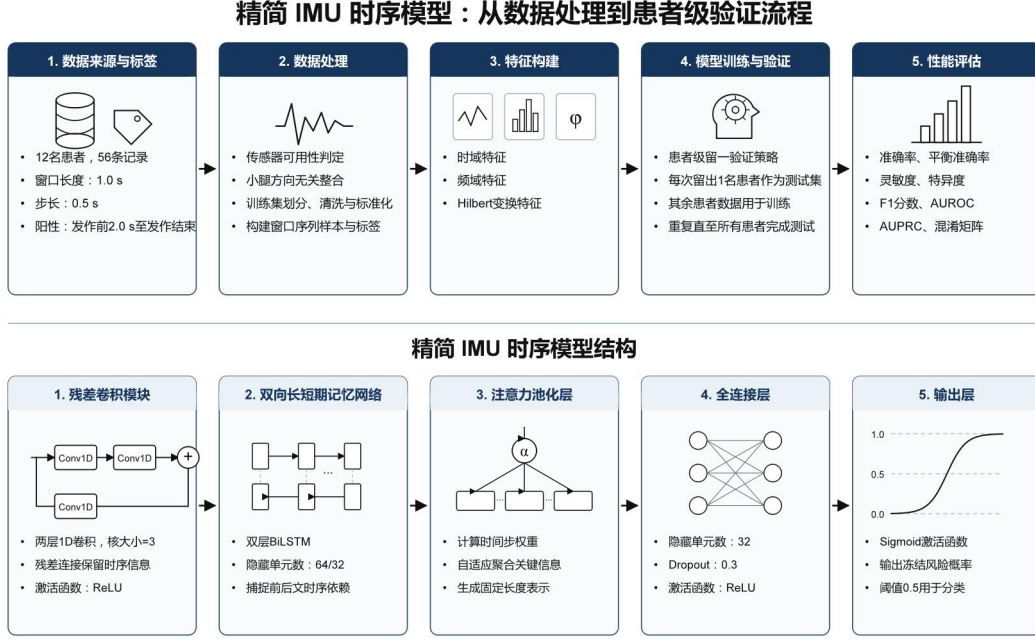


图 2 精简 IMU 时序模型及患者级验证流程

2.3 任务定义与无泄漏序列构建

原始信号采用长度为 1.0 s、步长为 0.5 s 的重叠窗口进行切分。以第 k 个窗口的结束时刻 e_k 作为预测时刻，连续 10 个窗口构成一个输入序列。由于相邻窗口重叠 0.5 s，一个序列的实际时间跨度为 $1.0 + 9 \times 0.5 = 5.5$ s。序列标签由末端窗口确定。设第 i 次 FOG 事件的起始和结束时刻分别为 t_i^{on} 和 t_i^{off} ，有效背景区间集合为 $\mathcal{B}$ ，则主二分类标签定义为：

$$y(e_k) = \begin{cases} 1, & t_i^{on} - 2.0 \leq e_k \leq t_i^{off}, \\ 0, & e_k \in \mathcal{B}, \\ \text{排除}, & t_i^{off} < e_k < t_i^{off} + 1.0. \end{cases} \quad (1)$$

为分析不同发作阶段，阳性样本进一步划分为 FOG 起始前 2.0 s 内的临近发作状态、窗口跨越 FOG 起始点的过渡状态、FOG 起始后 1.0 s 内的早期发作状态以及起始 1.0 s 后至发作结束的明确发作状态。FOG 结束后 1.0 s 定义为恢复灰区，不参与模型训练和主要评价。

2.4 可用性自适应小腿 IMU 表征与精简模型特征

2.4.1 采集记录传感器可用性

数据集中左、右小腿传感器的佩戴位置固定，但考虑到部分采集记录可能存在传感器无有效输出的情况，本研究使用采样频率为 500 Hz 的惯性信号，以每条记录前 2500 个采样点（即前 5.0 s 信号）作为判定依据，进行记录级传感器可用性判定。左、右小腿分别包含三轴加速度和三轴角速度共 6 个通道。当某侧小腿 6 个通道在检查区间内均为精确零值时，将

该侧判定为不可用；当各通道有限值比例均不低于 0.99，且至少一个加速度通道存在非零值时，将该侧判定为有效；其余情况标记为状态不明确，不纳入统一小腿表征的构建。传感器可用性判定结果在整条记录内保持固定，不根据单个窗口的运动幅度动态改变，以避免将静止或低幅度运动误判为传感器失效。仅一侧小腿传感器有效时，使用该信号提取特征；左右两侧均有效时，仅对不受传感器方向影响且两侧定义一致的同类特征进行均值汇聚，从而形成维度统一的统一小腿表征。设左右小腿对应的同类特征分别为 f_L 和 f_R ，则统一小腿特征 f_{shank} 定义为：

$$f_{\text{shank}} = \begin{cases} f_L, & \text{仅左侧有效,} \\ f_R, & \text{仅右侧有效,} \\ \frac{f_L + f_R}{2}, & \text{左右两侧均有效.} \end{cases} \quad (2)$$

对带符号的轴向均值、斜率及其他方向敏感特征，不进行左右侧均值汇聚，以避免相反方向的信号相互抵消。传感器可用性信息仅用于数据质量控制和特征汇聚，不作为模型输入，以降低模型利用设备故障模式识别患者身份的风险。

2.4.2 精简模型特征组与精简依据

精简模型 A2 的特征设计对应 3 类信号变化：运动幅度改变、步态与冻结频带之间的能量转移，以及窗口内的非平稳变化。为此，每个有效小腿窗口计算时域、频域、冻结—步态频带、Hilbert 包络和瞬时频率特征。前期开发性消融中，删除同位置加速度—角速度相位及同步特征后，主要指标未下降，部分指标反而改善，因此 A2 不再纳入该特征组。各特征组的代表量及作用见表 2。

表 2 精简模型特征组、信号含义及消融证据

特征组	代表特征	主要信息	实验依据
时域	均值、中位数、标准差、MAD、IQR、RMS、峰值、极差、波形长度、能量、加加速度 RMS	运动幅度、波动和冲击变化	B4 仅基础时频时性能下降
频域	主频、主频功率比、谱质心、谱带宽、谱熵、频带归一化功率	步态节律和频谱集中程度	与时域共同构成基础表示

特征组	代表特征	主要信息	实验依据
冻结一步态 频带	0.5—3.0 Hz 步态功率、3.0—8.0 Hz 冻结功率、功率比及谐波比	节律减慢、震颤样高频成分和步态冻结特征	删除后 BA 下降 0.0627
Hilbert 包络	包络均值、标准差、变异系数和斜率	幅值调制与肌动/运动强度变化	未见一致独立效应
瞬时频率	相位增量的中位数、IQR 和圆形统计量	短时频率波动和非平稳性	删除后轻度下降
未纳入精简方案	同位置加速度—角速度相位/同步特征	线性运动与旋转运动的相位耦合	删除后跨患者指标提高，形成 A2

对每个窗口计算均值、中位数、标准差、平均绝对偏差、中位绝对偏差、四分位距、均方根、绝对峰值、极差、波形长度、能量和加加速度均方根。这些量分别从幅值水平、离散程度和信号波动强度 3 个方面描述当前窗口。其中，设窗口内去均值信号为 x_n ，采样点数为 N ，其均方根和能量分别为：

$$RMS = \sqrt{\frac{1}{N} \sum_n x_n^2} \quad (3)$$

$$E = \sum_n x_n^2 \quad (4)$$

RMS 用于概括窗口内的总体信号幅值，能量反映该窗口运动强度的累积水平。二者数值下降可能对应肢体摆动减弱或运动停滞，但不能单独区分正常减速阶段与 FOG，因此还需结合频域及时序信息。

频域特征由去均值信号的快速傅里叶变换获得，包括主频、主频功率比例、频谱质心、频谱带宽、频谱熵、归一化频带功率和谐波功率比。根据原始特征定义，将 0.5~3.0 Hz 设为步态频带，将 3.0~8.0 Hz 设为冻结频带。设两频带功率分别为 P_{gait} 和 P_{freeze} ，则冻结与步态功率比为：

$$FI = \frac{P_{\text{freeze}}}{P_{\text{gait}} + \varepsilon} \quad (5)$$

式中， ε 为防止步态频带功率接近零而设置的极小正数。FI 表示 3.0~8.0 Hz 冻结频带功率与 0.5~3.0 Hz 步态频带功率的相对关系。冻结频带成分增多或步态频带成分减弱均会使 FI 升高。本研究仅把 FI 作为输入特征，不依据该数值单独确定 FOG 类别。

对带通滤波信号 $x(t)$ 进行 Hilbert 变换，并按式（6）构造解析信号。式（7）得到的包络用于计算窗口均值、标准差、变异系数和斜率。

$$z(t) = x(t) + j\mathcal{H}\{x(t)\} \quad (6)$$

$$a(t) = |z(t)| \quad (7)$$

$$\phi(t) = \text{unwrap}(\arg z(t)) \quad (8)$$

相位展开用于消除跨越 $[-\pi, \pi]$ 边界时产生的 2π 跳变，避免该跳变在差分后形成异常瞬时频率。其中， $\arg z(t)$ 表示解析信号的原始相位， $\text{unwrap}(x)$ 表示相位展开，即通过增加或减少 2π 消除由相位区间限制导致的突变，恢复相位随时间连续变化的趋势。若不进行该处理，相位跳变会在差分时产生异常大的瞬时频率。

设采样频率为 f_s ，瞬时频率按相邻采样点的展开相位计算：

$$f_{\text{inst}}(t) = \frac{f_s}{2\pi} [\phi(t) - \phi(t-1)] \quad (9)$$

初始特征方案还包含同一传感器位置加速度与角速度之间的相位差、相位锁定和同步特征，用于描述线性运动与旋转运动之间的耦合关系。前期同数据集开发性消融结果显示，删除该特征组后，主要性能指标未出现一致下降，部分指标反而有所改善，提示该特征组可能引入冗余信息或加剧患者间分布差异。因此，A2 精简方案保留时域、频域、冻结一步态频带、Hilbert 包络和瞬时频率特征，删除加速度—角速度相位及同步特征。

2.5 时序网络

2.5.1 残差卷积模块

输入序列首先经过 1 个残差卷积块。卷积分支包含两层一维卷积，每层使用 32 个长度为 3 的卷积核，并接批归一化。长度为 3 的卷积核覆盖相邻 3 个窗口，用于提取步态频带功率、包络和瞬时频率统计量的局部变化。当输入与卷积分支的通道数相同时，捷径分支直接传递输入；通道数不同时，使用一维卷积完成映射。两条分支相加后得到式（11）的残差输出。

$$H_1 = \text{ReLU}(\text{BN}(\text{Conv}_1(X))) \quad (10)$$

$$H_{\text{res}} = \text{ReLU}(\text{BN}(\text{Conv}_2(H_1)) + S(X)) \quad (11)$$

其中 $S(x)$ 为恒等捷径或投影捷径。残差连接通过恒等映射为局部卷积变换提供旁路，有助于缓解网络优化困难^[11]。本模型仅包含一个浅层残差块，其作用并非单纯加深网络，而是在局部特征变换后保留基础表示。

2.5.2 双层 BiLSTM

残差输出经随机失活层（Dropout, 0.3）后依次输入 64 单元和 32 单元的 BiLSTM^[12]，各层之后使用层归一化。两层网络均接收由 10 个历史窗口组成的序列。前向分支按时间顺序编码，反向分支从序列末端回看同一段历史，并按式（12）连接两路隐状态。由于 10 个窗口均位于预测时刻之前，该反向分支不包含预测时刻之后的数据。

$$h_t = [\vec{h}_t; \overleftarrow{h}_t] \quad (12)$$

前向分支描述从较早窗口向预测时刻的演化，反向分支则从序列末端回看同一段历史，以补充状态变化的上下文。由于全 10 个窗口都位于窗口预测时刻之前，双向编码不等于使用真实未来数据。B7 单向 LSTM 消融用于检验这种双向历史编码是否必要。

2.5.3 注意力池化与分类头

注意力池化使用全连接层（tanh 激活）给出标量得分^[13]，Softmax 权重后计算上下文向量：

$$e_t = \tanh(w^T h_t + b) \quad (13)$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_k \exp(e_k)} \quad (14)$$

$$c = \sum_t \alpha_t h_t \quad (15)$$

式（13）为每个时间位置计算标量得分，式（14）将得分转换为归一化权重，随后按式（15）对 10 个隐状态加权求和。权重根据各时间位置与当前分类任务的相关性分配^[13]，该权重仅用于模型内部的序列聚合，本文不把单个时间位置的高权重解释为临床生物标志物。最后依次连接含 32 个神经元的全连接层（ReLU 激活）、随机失活层（0.3）和 Sigmoid 分类头：

$$\hat{p} = \sigma(w_o^T \text{ReLU}(W_d c + b_d) + b_o) \quad (16)$$

$\hat{p}$ 表示当前序列属于 FOG 早期识别阳性类的概率，固定阈值 0.5 用于计算混淆矩阵、平衡准确率（Balanced Accuracy, BA）和 F1；受试者工作特征曲线下面积（area under the receiver operating characteristic curve, AUROC）与精确率-召回率曲线下面积（area under the precision-recall curve, AUPRC）由连续概率计算。

2.6 训练侧预处理与优化

每个外层折分别拟合一套预处理参数。非有限值填补、异常值截断界限、近零方差特征和相关性冗余特征均由训练患者确定，Z-score 的均值与标准差同样只由训练患者计算。验

证患者和测试患者仅应用这些参数，不参与重新估计，以避免患者间信息泄漏。

对于第 j 个特征，标准化过程表示为：

$$\tilde{x}_j = \frac{x_j - \mu_j^{train}}{\sigma_j^{train} + \varepsilon} \quad (17)$$

其中， μ_j^{train} 和 σ_j^{train} 分别表示训练患者对应特征的均值和标准差， ε 为防止除零的小常数。

由于 FOG 窗口与背景窗口存在类别不平衡，训练过程中根据训练序列类别比例计算类别权重，并采用加权二元交叉熵作为优化目标：

$$L = -\frac{1}{N} \sum_i [w_1 y_i \log(\hat{p}_i) + w_0 (1 - y_i) \log(1 - \hat{p}_i)] \quad (18)$$

其中， y_i 为真实标签， $\hat{p}_i$ 为模型预测概率， w_1 和 w_0 分别为正负类别权重。

训练使用 Adam 优化器，初始学习率设为 0.0005，每批包含 64 条序列，最多运行 10 轮。若验证损失连续 3 轮没有下降，则终止当前折训练，并恢复验证损失最低轮次对应的模型参数。

2.7 患者级验证、消融与统计分析

12 名患者依次承担外层测试对象，每一折完整留出其中 1 名患者，其余 11 名患者进入训练候选池。该设计评价的是同一数据集内对未见患者的内部泛化，而不是独立外部验证。对于第 s 个外层验证折，将患者 s 的全部数据作为测试集，其余患者组成训练候选池：

$$\mathcal{D}_{\text{test}}^{(s)} = \mathcal{D}_s, \mathcal{D}_{\text{pool}}^{(s)} = \bigcup_{k \neq s} \mathcal{D}_k \quad (19)$$

在每个外层患者留一折中，首先排除外层测试患者，再将训练候选池中的患者按照患者编号升序排列，选择编号最小且同时包含阳性和阴性序列的患者作为验证集，其余患者用于模型训练。若训练候选池中不存在同时包含两类序列的患者，则终止该外层折并记录错误。本研究各外层折均成功完成训练和测试。外层测试患者不参与缺失值处理、异常值截断、特征筛选、标准化参数估计、类别权重计算、早停、模型选择及分类阈值确定，以避免患者间信息泄漏。

精简 IMU 模型 A2 在消融实验中记为 B0，使用固定随机种子 20260718、20260719 和 20260720 重复训练，不根据结果选择最佳种子。B1—B4 分别删除 Hilbert、瞬时频率、冻结—步态频带特征或仅保留基础时频特征；B5—B7 分别去除残差、注意力或将双向 LSTM 替换为单向 LSTM；B8 增加 C3-Lite 因果动态特征；B9 在相同测试序列上融合表面肌电信号（sEMG）。个体化分析采用采集记录互斥的校准—测试划分，即用于个体化调整的记录不进入后续测试。P1 仅利用校准记录选择患者特异性分类阈值；P2 至 P4 则逐步扩大可训练

网络层的范围，用于比较阈值校准与不同程度参数微调的效果。跨患者分布差异采用标准化均值差、Wasserstein 距离、径向基函数核最大均值差异和域分类器曲线下面积进行描述。

统计分析以患者为基本单位。各消融实验首先计算候选模型相对于 A2 主模型的患者级指标差值，并报告患者级平均差。对于采用 3 个随机种子重复训练的残差消融，先在每位患者内计算多随机种子平均指标，再进行患者级配对比较，并采用 Bootstrap 法估计 95%置信区间。个体化分析报告平衡准确率的患者级平均差、95%置信区间及改善和退化患者数。分布偏移指标与模型性能之间的关系采用 Spearman 秩相关和置换检验分析，相关结果作为探索性证据解释。

3 结果

3.1 主模型 A2 的多随机种子表现

3 个随机种子均完成 12 个外层折，每个种子保存 23636 条测试序列的预测结果。患者宏平均 Accuracy、BA、F1、AUROC 和 AUPRC 分别为 0.7910 ± 0.0054 、 0.7378 ± 0.0122 、 0.6325 ± 0.0163 、 0.8791 ± 0.0058 和 0.7595 ± 0.0056 。AUROC 和 AUPRC 的标准差分别为 0.0058 和 0.0056，而 F1 的标准差为 0.0163，表明排序指标的种子间变化小于固定阈值分类结果。

表 3 A2 的三随机种子患者宏结果

随机种子	Accuracy	BA	F1	AUROC	AUPRC
20260718	0.785	0.751	0.622	0.885	0.764
20260719	0.794	0.735	0.625	0.873	0.753
20260720	0.794	0.727	0.651	0.879	0.761
均值±SD	0.791±	0.738±	0.633±	0.879±	0.760±
	0.005	0.012	0.016	0.006	0.006

注：宏平均按每项指标的可评价患者计算。BA 为平衡准确率；AUROC 为受试者工作特征曲线下面积；AUPRC 为精确率-召回率曲线下面积；SD 为标准差。

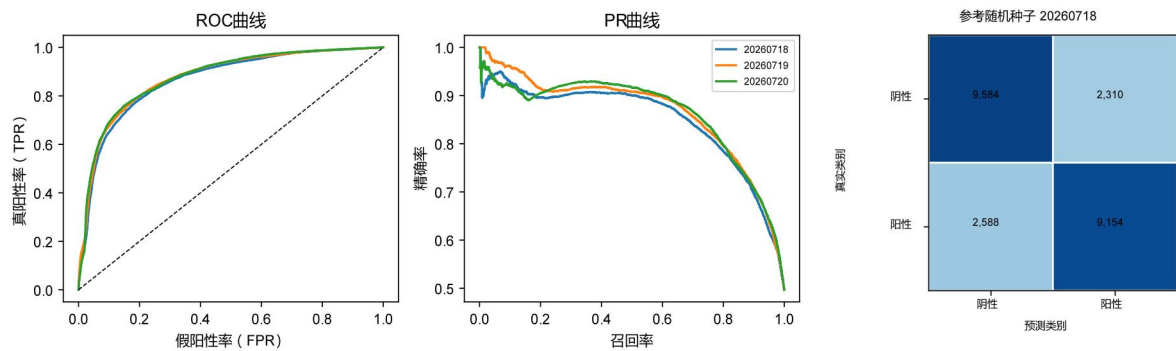


图 3 A2 的 ROC/PR 曲线与参考种子混淆矩阵

为进一步分析模型在不同 FOG 阶段的识别能力,将阳性序列按照其与 FOG 起始时刻的关系进行分层。3 个随机种子下, FOG 发作前 2.0 s、跨越 FOG 起始点、FOG 起 0~1.0 s 及明确 FOG 发作阶段的患者宏平均敏感度分别为 0.3321 ± 0.0155 、 0.5052 ± 0.0085 、 0.6088 ± 0.0118 和 0.8125 ± 0.0213 ; 各阶段与背景序列区分的患者宏平均 AUROC 分别为 0.6646 ± 0.0060 、 0.8046 ± 0.0052 、 0.8641 ± 0.0028 和 0.9307 ± 0.0012 。结果表明,模型的识别能力随 FOG 起始临近和发作进展逐步增强。发作前阶段已包含一定判别信息,但固定阈值下的敏感度较低,因此本研究结果支持 FOG 相关状态的早期识别,尚不能等同稳定的纯发作前预测。

参考种子 20260718 的混淆矩阵为 $TN=9584$ 、 $FP=2310$ 、 $FN=2588$ 、 $TP=9154$ 。患者间的性能差异较大(如患者 008 与患者 009 的 BA 相差 0.296),因此总体准确率不足以全面反映模型在患者级评价中的稳定性。患者 008 的平均 BA 和 AUROC 分别为 0.834 和 0.923;患者 009 分别为 0.538 和 0.759,平均敏感度仅 0.132、F1 为 0.194。对未包含阳性序列的患者 005,仍纳入准确率、特异度及阴性样本误分类情况的评价;其对应的敏感度、F1 值、AUROC 和 AUPRC 等依赖阳性样本的指标记为缺失值(NA)。患者级差异见表 4 和图 4。

表 4 A2 的每患者结果(3 个随机种子均值)

患者	Accuracy	BA	F1	AUROC	AUPRC
001	0.669	0.709	0.720	0.914	0.879
002	0.933	0.719	0.125	0.910	0.072
003	0.909	0.773	0.949	0.902	0.989
004	0.809	0.775	0.664	0.856	0.749
005	0.925	NA	NA	NA	NA
006	0.801	0.778	0.725	0.865	0.797

患者	Accuracy	BA	F1	AUROC	AUPRC
007	0.644	0.698	0.582	0.885	0.918
008	0.825	0.834	0.832	0.923	0.946
009	0.735	0.538	0.194	0.759	0.457
010	0.778	0.763	0.824	0.894	0.895
011	0.743	0.744	0.704	0.860	0.856
012	0.721	0.784	0.638	0.902	0.796

注：患者 005 无阳性序列，依赖正类或两类排序的指标记为 NA。

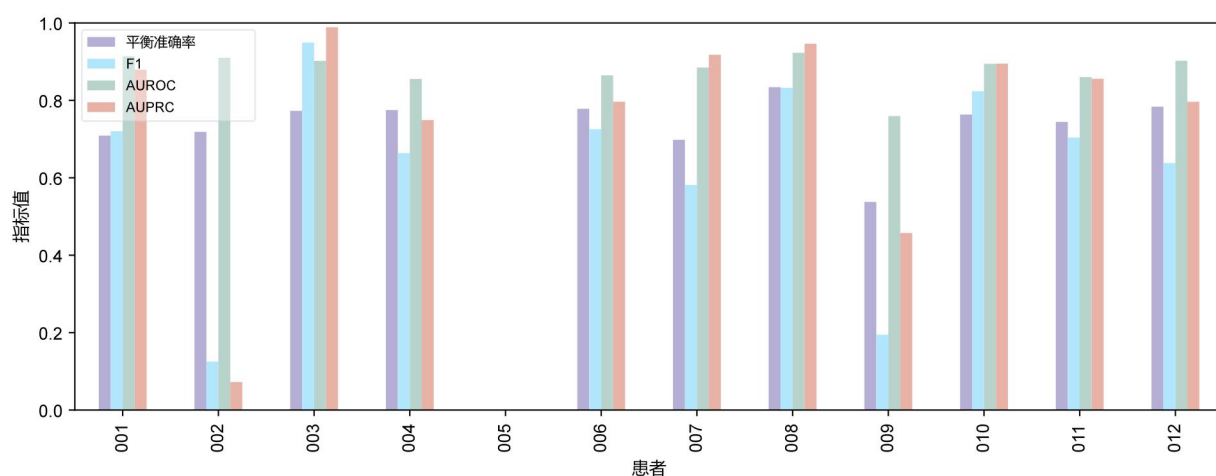


图 4 A2 各患者的主要性能指标

3.2 消融实验

3.2.1 特征消融

以 A2 参考种子为基线，去除冻结频带与步态频带功率比特征（B3）相对于 A2 的 BA、F1、AUROC 和 AUPRC 分别降低 0.0627、0.0410、0.0252 和 0.0270，是 4 组特征消融中下降最一致的方案。只有时域特征的模型（B4）的 4 项指标也均低于 A2。删除 Hilbert 包络后（B1），F1 提高 0.025，但 AUROC 和 AUPRC 分别降低 0.001 和 0.004；删除瞬时频率后（B2），4 项指标均略有下降。在参考随机种子 20260718 下，去除冻结频带与步态频带功率比后，多项指标呈较一致的下降趋势，提示该特征组可能是当前精简输入中的重要信息来源，Hilbert 和瞬时频率则主要提供互补信息。

3.2.2 网络模块消融

B7 将 BiLSTM 改为单向 LSTM 后，BA、F1、AUROC 和 AUPRC 分别降低 0.042、0.028、0.034 和 0.031，是结构消融中降幅最大的一组。B6 去除注意力池化后，4 项指标分别下降

0.007、0.009、0.006 和 0.012。残差模块使用 3 个随机种子重复比较, A2 相对于 B5 的 AUROC 和 AUPRC 平均提高 0.0053 和 0.0085, 但 BA 和 F1 的 95%置信区间跨越 0, 因此不将残差连接解释为对所有指标均有确定增益。

3.2.3 动态扩展、sEMG 融合与复杂度边界

C3-Lite (B8) 在 A2 基础上加入因果动态特征后, AUROC 和 AUPRC 分别变化-0.0005 和-0.0023, 未提高排序能力; F1 增加 0.0325, 但患者中位差为负, 提示提升由少数患者和固定阈值共同驱动。sEMG 融合(B9)与 B0 使用完全相同的 23636 条序列, BA 仅增加 0.0016, 而 AUROC 和 AUPRC 分别下降 0.0178 和 0.0200。结果不支持“更多动态特征或更多模态必然优于精简 IMU”的假设。各消融的方向汇总见表 5 和图 5。

表 5 A2 中心消融的患者级平均差

候选	AUPRC 差值	AUROC 差值	BA 差值	F1 差值
B1 去 Hilbert	-0.004	-0.001	-0.005	+0.025
B2 去瞬时频率	-0.008	-0.010	-0.016	-0.001
B3 去冻结一步态频带	-0.027	-0.025	-0.063	-0.041
B4 基础时频	-0.009	-0.013	-0.022	-0.001
B5 无残差	-0.008	-0.005	-0.001	-0.007
B6 无注意力	-0.012	-0.006	-0.007	-0.009
B7 单向 LSTM	-0.031	-0.034	-0.042	-0.028
B8 加 C3-Lite	-0.002	-0.001	-0.010	+0.033
B9 加 sEMG	-0.020	-0.018	+0.002	-0.006

注: 差值为候选模型指标减去 A2 主模型指标, 负值表示候选模型性能低于 A2。除去残差连接实验采用 3 个随机种子进行确认外, 其余消融结果均来自参考随机种子 20260718, 主要用于探索不同特征和结构的影响方向。

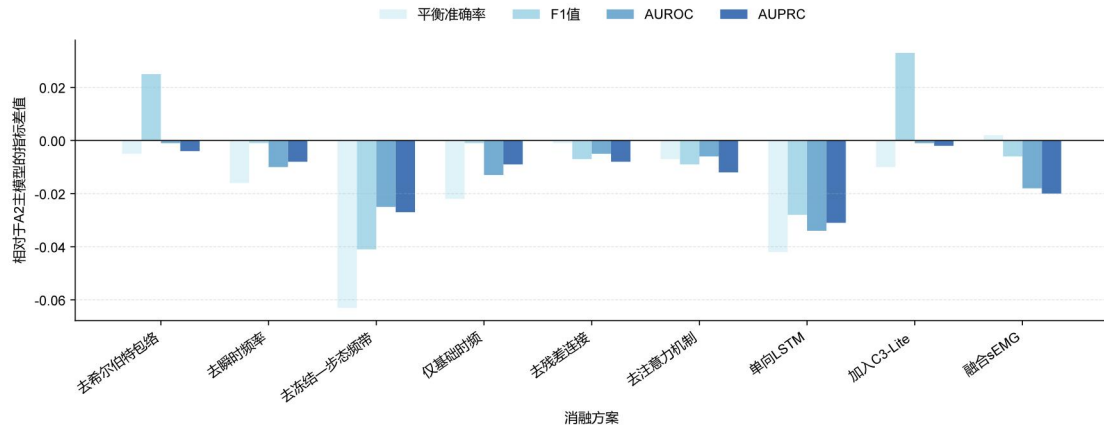


图 5 A2 中心特征、结构和模态消融结果

3.3 个体化校准与患者分布偏移

P1 在参与阈值校准的 10 名患者（排除患者 002 和 005）中的 BA 平均提高 0.0391，95% 置信区间为 0.0087~0.0683，其中 8 名患者改善、2 名退化。P1 只改变概率转换为类别标签时采用的阈值，不修改逐序列预测概率，因此校准前后的 AUROC 和 AUPRC 数值相同。P2、P3 和 P4 的 BA 平均变化分别为 0.002、0.007 和 0.003，且患者间方向不一致。现有结果支持优先调整分类阈值，尚不足以说明扩大参数微调范围能够稳定改善跨患者结果。

在全部 11 名可评价患者中（排除无阳性样本的患者 005），平均 Wasserstein 距离与 BA 的 Spearman 相关系数为-0.800，置换检验 $P = 0.005$ ；径向基函数核 MMD 与 BA 的相关系数为-0.818， $P = 0.004$ 。患者 009 的 MMD 最高，但其他 3 项偏移指标并非全部最高。因此，该患者的低敏感度不能仅用整体分布距离解释，类别条件分布或多个特征的联合偏移也可能共同作用。

表 6 个体化策略平衡准确率的变化

策略	患者数	BA 均值差	95%置信区间	改善/退化/不变
P1 阈值校准	10	+0.039	+0.009~+0.068	8/2/0
P2 仅分类头	10	+0.002	+0.000~+0.004	5/1/4
P3 部分微调	10	+0.007	-0.002~+0.016	7/3/0
P4 较强微调	10	+0.003	-0.001~+0.008	7/3/0

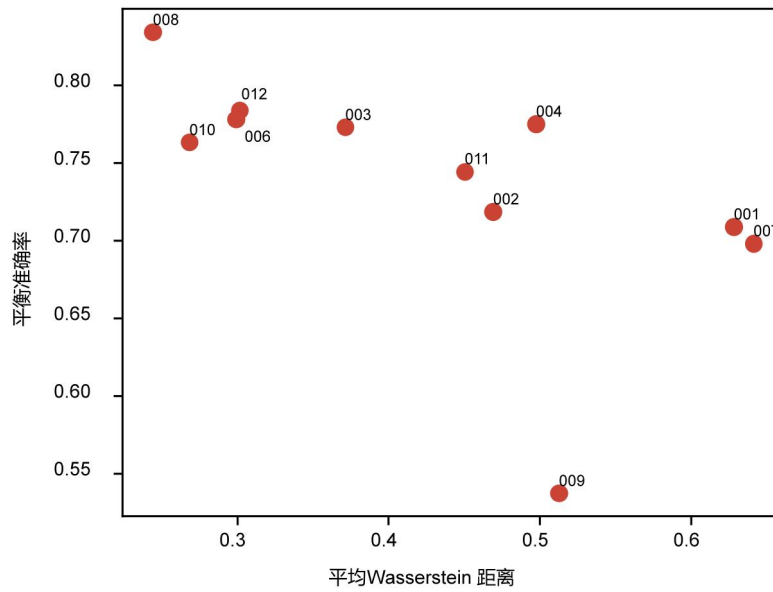


图 6 分布偏移分析

注：分布偏移分析纳入了全部可计算 BA 的 11 名患者，与阈值校准的纳入范围不同。

3.4 困难患者 009 的失败模式

患者 009 在固定阈值 0.5 下主要表现为漏检，而不是误报：参考随机种子结果为 TN=1028、FP=48、FN=363、TP=9，特异度为 0.955，敏感度为 0.024。3 个随机种子验证中，该患者的 AUROC 为 0.7305~0.7849，但仍处于低敏感度和低 F1 值区间。其 6 条采集记录的预测概率中位数整体偏低，FOG 发作前 2.0 s 阶段几乎未被检出；双侧小腿传感器均有效，缺失率为 0，异常值截断比例也较低。该病例提示，模型可能学到了总体可排序的风险，但在部分患者中仍存在预测概率整体偏低和正类表征不足的问题。

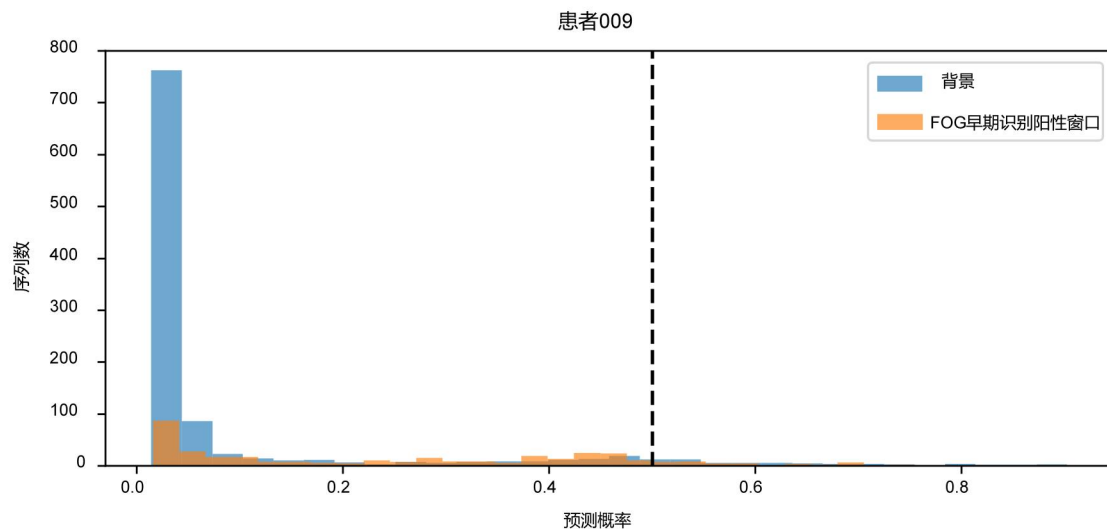


图 7 患者 009 的背景与 FOG 相关序列预测概率分布

4 讨论

4.1 主要发现与创新

记录级传感器判定解决了部分采集记录中单侧小腿信号无效的问题。单侧有效时直接使用有效侧，双侧有效时只汇聚方向无关的同类特征；传感器可用性状态没有作为模型输入。这样可以保持输入维度一致，同时减少模型利用设备缺失模式区分患者的可能。A2 删除加速度—角速度相位及同步特征也有消融依据，而不是单纯压缩特征数量。

早期 FOG 监测研究已采用步态低频带与冻结高频带之间的能量关系描述步态冻结特征^[14]。后续研究发现，传感器类型、佩戴位置、活动任务及数据来源都会影响模型结果，假报警问题在外部应用场景中仍较突出^[15]。本研究的消融结果与上述认识一致：删除冻结—步态频带特征后，平衡准确率、AUROC 和 AUPRC 分别下降 0.0627、0.0252 和 0.0270；将 BiLSTM 替换为单向 LSTM 后，上述 3 项指标分别下降 0.042、0.034 和 0.031。由此可见，频带能量关系和 10 个历史窗口内的双向时序信息是当前模型的重要输入。然而，上述特征的有效性并非在所有条件下都稳定。已有研究还指出，活动任务和用药状态会改变 FOG 模型的识别表现^[16]，这意味着同一患者的步态信号在不同情境下可能存在较大差异。这种条件依赖性进一步提示，模型的性能不能仅以全部序列的总体指标衡量，还需检验其面对不同患者时的真实泛化能力。为此，本研究采用患者级留一验证框架，每个外层折完整隔离一名患者的全部数据，预处理参数、早停及模型选择均不接触测试患者，以更严格地暴露模型在面对不同患者个体特征时的性能差异。

在此评价框架下，本研究发现增加因果动态特征、融合表面肌电信号或扩大个体化微调范围均未形成稳定增益；相反，仅调整分类阈值（P1）便使 10 名参与校准的患者中有 8 名的平衡准确率提高。这一结果表明，当前数据条件下的患者差异更多体现在分类边界，而深层参数微调是否有效仍缺乏稳定证据。表面肌电可用于帕金森病步态事件预测^[17]，肌电与惯性信号的联合分析也已用于 FOG 研究^[18]；但本研究在相同测试序列上的比较表明，模态互补性需在通道对齐、样本匹配和融合策略均合理设计的前提下，才可能转化为跨患者收益。整体而言，少量校准数据更适合用于修正患者间的决策边界，而非稳定重塑深层时序表示。

本文完整记录了患者划分、训练侧预处理、随机种子和评价过程，符合 TRIPOD+AI 强调的透明报告原则^[19]。按照 PROBAST+AI 的评价框架^[20]，当前研究最主要的风险来自样本量较小，以及特征开发与最终评价均使用同一数据集。国内医学信息学研究也同样指出，数据来源和验证方式会影响机器学习模型的可迁移性^[21]；面向老年人群的人工智能筛查研究

进一步强调了人群异质性和真实场景评价的重要性^[22]。近期跨数据集研究进一步表明，单一数据集内的较高性能并不必然转化为新患者和新场景中的稳定表现^[23]。因此，本研究结果应解释为患者级内部验证证据，不能替代独立外部验证。

4.2 局限性与展望

本研究仅纳入 12 名患者，A2 特征方案与最终 LOSO 评价均来自同一公开数据集，可能存在模型选择乐观偏倚。患者之间的标签构成也不均衡：患者 005 没有阳性序列，患者 002 仅有 10 条阳性序列，其 F1 值和 AUPRC 容易受到少量预测结果影响。此外，阳性标签覆盖 FOG 发生前 2.0 s 至发作结束，因此本文评价的是 FOG 相关状态的早期识别，而非严格意义上的纯发作前预测。

此外，本研究采用窗口级分类评价，未在连续时间轴上统计单位时间假报警次数，该指标有待后续在连续预警场景中进一步验证。最后，个体化校准记录的顺序尚不能确认为真实时间顺序，分布偏移与性能的相关分析亦受患者数量限制，只能作探索性解释。后续研究应进一步检验模型的迁移能力与临床应用边界^[23]。

5 结论

本研究完成了精简 IMU 时序模型的患者级留一内部验证。模型以记录级传感器可用性判定和方向无关汇聚构建统一小腿输入，再由残差一维卷积、双向 LSTM 和注意力池化处理 10 个连续窗口。3 个固定随机种子下，患者宏平均 AUROC 为 0.8791，AUPRC 为 0.7595。参考随机种子的消融结果显示，删除冻结一步态频带特征或采用单向 LSTM 后，多项指标下降；C3-Lite 动态扩展和 sEMG 融合则没有改善 AUROC 和 AUPRC。与继续叠加特征、模态或深层个体化微调相比，该模型在当前数据条件下表现出更好的简洁性、可复现性和跨患者稳定性，可为 FOG 患者级早期识别及异质性分析提供方法学参考与模型基准，但其临床价值仍需独立队列和连续报警场景中进一步验证。

作者贡献：赵乙璇负责研究设计、算法实现和论文撰写；高瑞负责数据预处理、结果核对及模型性能分析；任依飒负责数据收集和统计；肖启睿负责数据处理的程序编写；王玉锋负责论文写作的学术规范；高建强负责研究方案指导和论文审阅；李莉负责论文写作质量控制和论文修订。

利益声明：所有作者均声明不存在利益冲突。

参考文献

- [1] NUTT J G, BLOEM B R, GILADI N, et al. Freezing of gait: moving forward on a mysterious clinical phenomenon[J]. *The Lancet Neurology*, 2011, 10(8):734-744.
- [2] SINGHAL M, GRANNIE C, BURNETTE M, et al. Freezing of gait in Parkinson's disease: a scoping review on the path towards real-time therapies[J]. *Sensors*, 2026, 26(7): 2042.
- [3] ZHANG W, SUN H, HUANG D, et al. Detection and prediction of freezing of gait with wearable sensors in Parkinson's disease[J]. *Neurological Sciences*, 2024, 45(2): 431-453.
- [4] PALMERINI L, ROCCHI L, MAZILU S, et al. Identification of characteristic motor patterns preceding freezing of gait in Parkinson's disease using wearable sensors[J]. *Frontiers in Neurology*, 2017, 8:394.
- [5] HAUSSLER A M, TUETH L E, MAY D S, et al. Refinement of an algorithm to detect and predict freezing of gait in Parkinson disease using wearable sensors[J]. *Sensors*, 2025, 25(1): 124.
- [6] FU Z, LIN C, ZHANG Y. Personalized prediction of gait freezing using dynamic mode decomposition[J]. *Scientific Reports*, 2025, 15: 18749.
- [7] ABBASI S, REZAAEE K. Deep learning-based prediction of freezing of gait in Parkinson's disease with the ensemble channel selection approach[J]. *Brain and Behavior*, 2025, 15(1): e70206.
- [8] WANG B, HU X, GE R, et al. Prediction of freezing of gait in Parkinson's disease based on multi-channel time-series neural network[J]. *Artificial Intelligence in Medicine*, 2024, 154: 102932.
- [9] KLAVER E C, HEIJINK I B, SILVESTRI G, et al. Comparison of state-of-the-art deep learning architectures for detection of freezing of gait in Parkinson's disease[J]. *Frontiers in Neurology*, 2023, 14: 1306129.
- [10] ZHANG W, YANG Z, LI H, et al. Multimodal data for the detection of freezing of gait in Parkinson's disease[J]. *Scientific Data*, 2022, 9: 606.
- [11] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770-778.
- [12] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [13] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[C] //Proceedings of the 3rd International Conference on Learning Representations. San Diego, 2015.
- [14] MOORE S T, MACDOUGALL H G, ONDO W G. Ambulatory monitoring of freezing of gait

- in Parkinson's disease[J]. *Journal of Neuroscience Methods*, 2008, 167(2): 340-348.
- [15] BORZÌ L, DEMROZI F, BACCHIN R A, et al. Freezing of gait detection: the effect of sensor type, position, activities, datasets, and machine learning model[J]. *Journal of Parkinson's Disease*, 2025, 15(1): 163-181.
- [16] YANG P K, FILTJENS B, GINIS P, et al. Freezing of gait assessment with inertial measurement units and deep learning: effect of tasks, medication states, and stops[J]. *Journal of NeuroEngineering and Rehabilitation*, 2024, 21(1): 24.
- [17] HAUFE S, ISAIAS I U, PELLEGRINI F, et al. Gait event prediction using surface electromyography in parkinsonian patients[J]. *Bioengineering*, 2023, 10(2): 212.
- [18] MAZZETTA I, ZAMPOGNA A, SUPPA A, et al. Wearable sensors system for an improved analysis of freezing of gait in Parkinson's disease using electromyography and inertial signals[J]. *Sensors*, 2019, 19(4): 948.
- [19] COLLINS G S, MOONS K G M, DHIMAN P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods[J]. *BMJ*, 2024, 385: e078378.
- [20] MOONS K G M, DAMEN J A A, KAUL T, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods[J]. *BMJ*, 2025, 388: e082505.
- [21] 任媛渊, 丁福, 付荣娟. 国外基于机器学习的住院患者跌倒风险预测模型构建研究及其启示[J]. *医学信息学杂志*, 2023, 44(10): 63-67, 80.
- [22] 司唯, 徐婷, 林家玥, 等. 人工智能技术在老年认知症筛查中的应用现状[J]. *医学信息学杂志*, 2025, 46(2): 36-41, 47.
- [23] SALOMON A, GAZIT E, SHIMONI N, et al. Beyond the leaderboard: evaluating the robustness of deep learning models for detecting freezing of gait[J]. *European Journal of Neuroscience*, 2026, 63(3): e70424.