

doi: 10.3969/j.issn.1674-1242.2026.04.002

基于改进型 BiLSTM 的多参数监护仪预防性维护预测模型的构建与验证

范璐敏^{1,2}, 马萱航², 樊重俊^{2*}

(1. 同济大学附属东方医院 资产管理部, 上海 200120; 2. 上海理工大学 管理学院, 上海 200093)

【摘要】目的 构建引入时序注意力机制的改进型双向长短期记忆网络(bidirectional long short-term memory, BiLSTM)模型,用于多参数监护仪的故障预警。**方法** 收集同济大学附属东方医院急诊科 96 台监护仪的质控检测数据及维修记录,以功能通道为建模单元,采用滑动窗口法生成 5 674 个通道级时序样本,按 7:3 比例划分训练集与测试集。在 BiLSTM 编码层后引入时序注意力机制,以长短期记忆网络(long short-term memory, LSTM)、门控循环单元(gated recurrent unit, GRU)、时序卷积网络(temporal convolutional network, TCN)及标准 BiLSTM 作为参比模型。综合受试者操作特征(receiver operator characteristic, ROC)曲线的曲线下面积(area under the curve, AUC)、精确率-召回率(precision-recall, PR)曲线、概率校准曲线及阈值下分类指标对各模型进行系统评估,筛选综合性能最优模型。**结果** 改进型 BiLSTM 综合性能最优, AUC 为 0.918, 平均精确率(average precision, AP)为 0.553, 敏感度为 80.43%, F_1 评分为 65.49%, 与各参比模型的 AUC 差异均有统计学意义(均 $P < 0.05$)。Hosmer-Lemeshow 检验提示模型校准良好($P = 0.480$), 其余参比模型均存在明显校准偏差(均 $P < 0.001$)。**结论** 改进型 BiLSTM 模型能识别多参数监护仪质控参数的跨周期退化趋势,在故障预警和概率校准方面均优于参比模型,可为医疗设备预防性维护提供辅助决策依据。

【关键词】 多参数监护仪; 预防性维护; 双向长短期记忆网络; 时序注意力机制**【中图分类号】** R197.39**【文献标志码】** A

文章编号: 1674-1242 (2026) 04-0010-06

Construction and validation of a predictive model for preventive maintenance of multi-parameter patient monitor based on an improved BiLSTM

FAN Lumin^{1,2}, MA Xuanhang², FAN Chongjun^{2*}

(1. Asset Management Department, Shanghai East Hospital, School of Medicine, Tongji University, Shanghai 200120, China;

2. Business School, University of Shanghai for Science and Technology, Shanghai 200093, China)

【Abstract】Objective To construct an improved bidirectional long short-term memory (BiLSTM) model incorporating a temporal attention mechanism for fault warning of multi-parameter patient monitor. **Methods** Quality control records and maintenance data from 96 patient monitors in the emergency department of Shanghai East Hospital, Tongji University, were collected. Functional channels were used as modeling units, and 5,674 channel-level time-series samples were generated using a sliding-window approach. The samples were divided into training and test sets at a ratio of 7 vs. 3. A temporal attention layer was added after the BiLSTM encoder, and long short-term memory (LSTM), gated recurrent unit (GRN), temporal convolutional network (TCN), and standard BiLSTM models were used as comparators. The models were systematically evaluated using the area under the curve (AUC) of receiver operator characteristic (ROC) curve, precision-recall (PR) curve, probability calibration curve, and threshold-dependent classification metrics to identify the model with the best overall performance. **Results** The improved BiLSTM achieved the best overall performance, with an AUC of 0.918, average precision (AP) of 0.553, sensitivity of 80.43%, and F_1 score of 65.49%. The AUC differences between the improved BiLSTM and all comparator models were statistically significant (all $P < 0.05$). The Hosmer-Lemeshow test indicated good calibration of the improved model ($P = 0.480$), whereas the comparator models showed significant calibration bias (all $P < 0.001$). **Conclusion** The improved BiLSTM model can identify cross-cycle degradation trends in quality control parameters of multi-parameter

收稿日期: 2026-01-22。

基金项目: 上海市中西医结合学会医学工程专业委员会医学工程科研基金项目 (2025-YG-013)。

作者简介: 范璐敏, 博士研究生, 研究方向: 智慧医疗管理。E-mail: 221420093@st.usst.edu.cn。

通信作者: 樊重俊, 博士, 教授, 研究方向: 智慧医疗、深度学习教学与科研。E-mail: cjfan@usst.edu.cn。

patient monitor. It outperforms comparator models in both fault warning and probability calibration, and can provide auxiliary decision-making basis for preventive maintenance of medical equipment.

【Key words】 Multi-parameter patient monitor; Preventive maintenance; Bidirectional long short-term memory; Temporal attention mechanism

多参数监护仪是急诊科用于生命体征监测的核心医疗设备,其运行质量直接关乎急危重症患者的救治安全^[1]。在医疗设备维护管理中,提升维护效率、管控运行风险已成为医院精细化管理的重要组成部分^[2]。目前,医疗机构多采用阈值判定的被动维护模式,仅依据单次质控参数是否超出允许误差范围判定维护需求,难以识别参数尚处于合格区间内的渐进性退化趋势^[3]。近年来,预测性维护和机器学习方法已逐步应用于医疗设备故障预测领域,但现有研究大多依托设备静态特征、维修记录或实时传感数据展开建模^[4-5],对常规质控参数的跨周期时序退化规律挖掘相对不足。双向长短期记忆网络(bidirectional long short-term memory, BiLSTM)能捕捉时序数据的双向依赖特征,在故障预测任务中具备较好适用性^[6];引入时序注意力机制后,模型可进一步强化对关键时间步的识别能力^[7-8]。但上述方法在监护仪定期质控数据故障预测场景的应用尚属空白。基于此,本研究通过构建引入时序注意力机制的改进型 BiLSTM 预防性维护预测模型,以期实现设备故障主动预警,为医学工程部门开展前瞻性维护提供依据。

1 资料与方法

1.1 数据集

1.1.1 数据来源

本研究收集 2021 年 1 月至 2025 年 12 月同济大学附属东方医院急诊科 96 台在用多参数监护仪的季度质控检测数据及维修记录。所有检测数据均与设备资产编号对应,可追溯各设备及功能通道的历次检测轨迹。纳入设备涵盖迈瑞、飞利浦及 GE 等临床常用品牌,设备在用年限组间差异无统计学意义($P>0.05$)。

1.1.2 研究单元定义与数据采集

以功能通道作为建模单元。每台监护仪包括心率(heart rate, HR)、呼吸率(respiratory rate, RR)、血压(blood pressure, BP)及外周血氧饱和度(peripheral oxygen saturation, SpO₂)4类功能通道,96台设备共形成 384 条通道级质控时间序列。各通道序列由历次

质控检测结果按时间顺序排列构成。HR、RR 及 SpO₂通道分别记录多个标准设定点的测量误差,BP 通道合并收缩压(systolic blood pressure, SBP)和舒张压(diastolic blood pressure, DBP)误差进行建模。所有特征值均记录为测量值与设定值之差的整数形式。

1.1.3 样本构建方法

采用滑动窗口法对各通道时序序列进行样本提取,窗口长度设定为 4 个连续质控时间步($W=4$),以步长 1 逐步滑动截取,每个连续子序列构成一个独立窗口样本,序列长度不足 W 者不予纳入。 $W=4$ 既满足捕捉非单调趋势变化的时序最低长度要求($W\geq 3$),也与每季度执行一次质控的制度对应,涵盖一个完整质控年度的时序跨度。当某通道发生维修事件时,维修完成后的质控记录视为新时序段的起点,原时序段在维修事件处截断。样本标注以窗口末端时间步为基准:质控检出故障(即第 n 次质控发现参数异常),以 t_{n-1} 为末端的窗口标注为故障预警样本,末端早于 t_{n-1} 的窗口标注为正常样本;质控间期故障(即第 n 次质控通过后、第 $n+1$ 次质控前突发故障),以 t_n 为末端的窗口标注为故障预警样本,末端早于 t_n 的窗口标注为正常样本。观察期内无故障记录的通道,其全部窗口样本均标注为正常样本。

1.2 数据预处理

按设备编号将 96 台监护仪以 7:3 比例随机分为训练集和测试集,同一设备的全部通道样本归入同一数据集。缺失值采用多重插补法处理,连续变量以训练集均值和标准差进行 Z-score 标准化,再将相同参数应用于测试集。针对训练集中类别不平衡问题,采用 SMOTE 对故障预警样本进行过采样,测试集保持原始类别分布。

1.3 模型构建

1.3.1 BiLSTM 基础架构

BiLSTM 模型以形状为 $W\times F$ 的通道级质控时序矩阵作为输入,其中 W 为窗口长度, F 为通道特征维度。输入序列经正向和反向 LSTM 编码后拼接形成时序特征表示,最终经 Sigmoid 层输出下一质控周期

前发生故障的预测概率。

1.3.2 时序注意力机制改进

标准 BiLSTM 对各时间步隐藏状态采用等权重聚合,无法反映不同质控周期对故障预测的差异性贡献。为此,在 BiLSTM 编码层之后引入时序注意力层:BiLSTM 对输入序列编码后,在第 t 个时间步输出隐藏状态向量 $h_t \in R^{2u}$ (u 为单向 LSTM 隐藏单元数, R^{2u} 表示双向拼接后所在的 $2u$ 维实数向量空间),全部时间步的隐藏状态构成序列 $H = \{h_1, h_2, \dots, h_W\}$ 。注意力层通过可学习权重矩阵 W_a 与偏置向量 b_a , 经 $e_t = \tanh(W_a h_t + b_a)$ 计算各时间步的非归一化注意力得分 e_t , 再经 Softmax 函数归一化得到各时间步的注意力权重 $a_t = \exp(e_t) / \sum_{j=1}^W \exp(e_j)$, 以加权求和 $c = \sum_{t=1}^W a_t h_t$ 得到上下文向量 c 并输入输出层。 W_a 和 b_a 在模型训练过程中端到端更新,无需人为预设权重规则。

1.3.3 模型训练方案

模型采用 Adam 优化器和二元交叉熵损失函数开展训练,初始学习率设为 0.001,结合余弦退火、Dropout、L2 正则化及早停策略控制过拟合。通过网格搜索联合五折交叉验证确定隐藏单元数、网络层数及 L2 正则化系数,最终在完整训练集上重新训练最优模型,于独立测试集完成性能评估。

1.4 对比实验与评估指标

1.4.1 对比实验

选取长短期记忆网络(long short-term memory, LSTM)、门控循环单元(gated recurrent unit, GRU)、时序卷积网络(temporal convolutional network, TCN)及标准 BiLSTM 作为对比模型。所有模型均采用相同的输入样本、训练集和测试集划分规则,训练策略保持一致。其中标准 BiLSTM 与主模型结构基本一致,仅移除时序注意力层,用于单独评估注意力机制的增量贡献。

1.4.2 评估指标

绘制受试者操作特征(receiver operator characteristic, ROC)曲线并计算曲线下面积(area under the curve, AUC),评价模型区分故障预警样本与正常样本的整体判别能力;绘制精确率-召回率(precision-recall, PR)曲线并计算平均精确率(average precision, AP),评估类别不平衡条件下模型对少数类样本的识别性能;采用校准曲线分析预测

概率与实际故障发生率的一致性;依据 ROC 曲线约登指数确定最优决策阈值,并计算敏感度、特异度、精确率、准确率及 F_1 评分。

1.5 统计学方法

采用 Python(v3.12.0)软件完成统计学分析与模型构建。符合正态分布的计量资料以均数 $\pm$ 标准差($\bar{x} \pm s$)表示,组间比较采用 t 检验。模型校准性能采用 Hosmer-Lemeshow 检验评估,各性能指标的 95% 置信区间采用 Bootstrap 法(重抽样 1 000 次)估计,模型间 AUC 比较采用 DeLong 检验。 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 数据集特征与预处理结果

结果显示,故障预警窗口的测量误差随质控周期推进呈上升趋势,而正常窗口误差波动相对稳定(图 1)。采用滑动窗口法共生成 5 674 个窗口样本,其中训练集 3 970 例,测试集 1 704 例。训练集与测试集故障预警样本占比分别为 5.0% 和 5.4%,差异无统计学意义($P > 0.05$)。两组 25 项通道设定点误差中仅 4 项差异有统计学意义(均 $P < 0.05$),且呈散在分布(表 1)。SMOTE 预处理后,训练集中各通道故障预警样本量分别由原始的 52、50、46、52 例扩充至 319、318、319、322 例,以优化模型训练效能,测试集保持原始类别分布。

2.2 模型训练与超参数优化结果

网格搜索联合五折交叉验证结果显示,改进型 BiLSTM 在隐藏单元数为 64、网络层数为 1 层、L2 正则化系数为 1×10^{-4} 时获得最高平均 F_1 。最终模型训练参数设定为初始学习率 0.001、Dropout 比率 0.2、批量大小 32。在最优超参数配置下采用完整训练集重新训练模型,模型于第 42 个 epoch 触发早停机制并结束训练,训练损失由 0.562 持续下降至 0.148 后趋于稳定。

2.3 模型对比实验

改进型 BiLSTM 的 AUC 为 0.918, AP 为 0.553,均高于标准 BiLSTM、LSTM、GRU 和 TCN,且 AUC 差异均具有统计学意义(均 $P < 0.05$)(图 2A-B)。校准分析显示,改进型 BiLSTM 的检验统计量为 7.535($P = 0.480$),整体概率校准性能良好,但在预测概率较高区间,曲线存在偏离完美校准线的趋势;其余 4 个模型的检验统计量均显著升高(均 $P < 0.001$)(图 2C)。

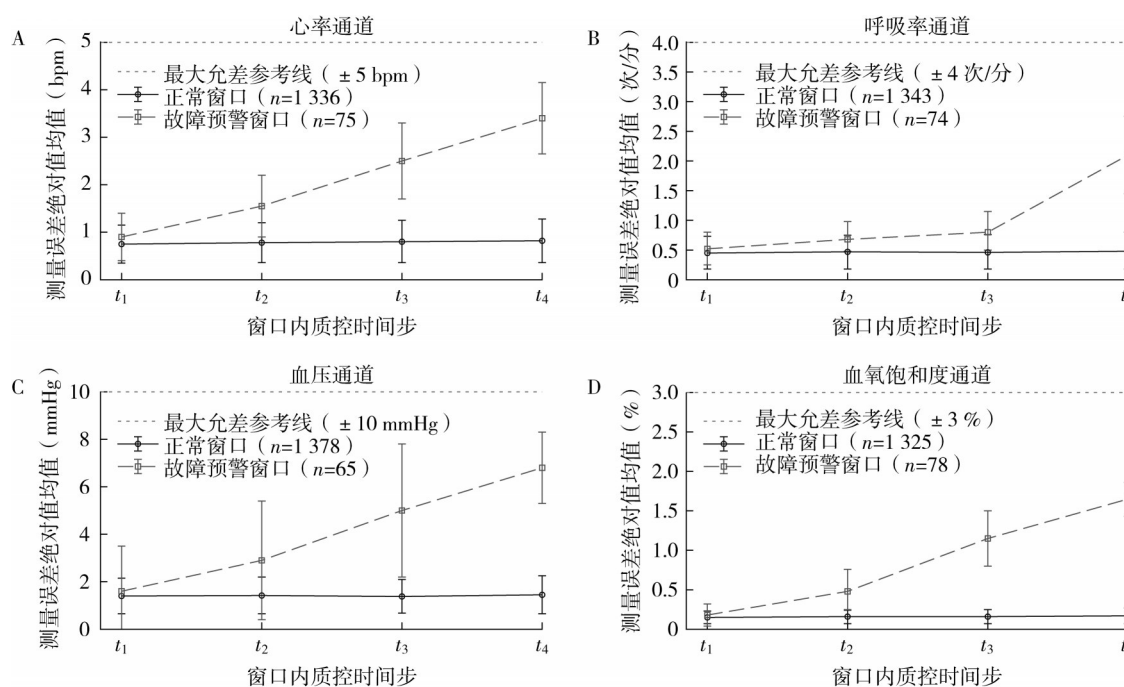


图1 不同通道测量误差时序变化趋势

表1 不同通道参数绝对测量误差比较($\bar{x} \pm s$)

通道	参数	设定点	训练集 ($n=3970$)	测试集 ($n=1704$)	t	P
HR	HR (次/分)	30	1.01 $\pm$ 0.96	1.02 $\pm$ 1.03	-0.162	0.871
		60	1.09 $\pm$ 0.98	1.09 $\pm$ 0.96	0.008	0.994
		100	1.09 $\pm$ 0.96	1.13 $\pm$ 1.01	-1.111	0.267
		120	1.13 $\pm$ 0.99	1.14 $\pm$ 1.01	-0.363	0.716
		180	1.10 $\pm$ 0.98	1.07 $\pm$ 1.00	1.127	0.260
RR	RR (次/分)	15	0.58 $\pm$ 0.69	0.64 $\pm$ 0.70	-2.615	0.009
		20	0.61 $\pm$ 0.69	0.63 $\pm$ 0.67	-1.405	0.116
		40	0.58 $\pm$ 0.68	0.61 $\pm$ 0.68	-1.397	0.163
		60	0.60 $\pm$ 0.67	0.58 $\pm$ 0.67	1.045	0.296
		80	0.59 $\pm$ 0.67	0.61 $\pm$ 0.69	-0.905	0.366
BP	SBP (mmHg)	60	1.82 $\pm$ 1.62	1.90 $\pm$ 1.66	-1.823	0.068
		80	1.77 $\pm$ 1.60	1.80 $\pm$ 1.58	-1.140	0.132
		100	1.87 $\pm$ 1.57	1.73 $\pm$ 1.57	3.128	0.002
		120	1.82 $\pm$ 1.55	1.88 $\pm$ 1.59	-1.168	0.243
		150	1.81 $\pm$ 1.63	1.81 $\pm$ 1.55	0.004	0.997
	DBP (mmHg)	30	1.79 $\pm$ 1.53	1.90 $\pm$ 1.59	-2.525	0.012
		50	1.81 $\pm$ 1.54	1.79 $\pm$ 1.56	0.398	0.691
		65	1.79 $\pm$ 1.56	1.86 $\pm$ 1.60	-1.406	0.160
		80	1.82 $\pm$ 1.61	1.75 $\pm$ 1.52	1.645	0.100
		100	1.89 $\pm$ 1.63	1.78 $\pm$ 1.58	2.487	0.013
SpO ₂	SpO ₂ (%)	85	0.23 $\pm$ 0.48	0.24 $\pm$ 0.51	-0.384	0.701
		88	0.23 $\pm$ 0.48	0.21 $\pm$ 0.49	1.443	0.149
		95	0.26 $\pm$ 0.50	0.24 $\pm$ 0.50	1.469	0.142
		98	0.26 $\pm$ 0.49	0.27 $\pm$ 0.53	-1.231	0.218
		100	0.23 $\pm$ 0.48	0.26 $\pm$ 0.50	-1.939	0.053

注: HR: 心率; RR: 呼吸率; BP: 血压; SBP: 收缩压; DBP: 舒张压; SpO₂: 外周血氧饱和度。

在最优决策阈值下,改进型 BiLSTM 在敏感度、特异度、精确率、准确率及 F_1 评分各项分类指标上均取得最优表现(表2)。

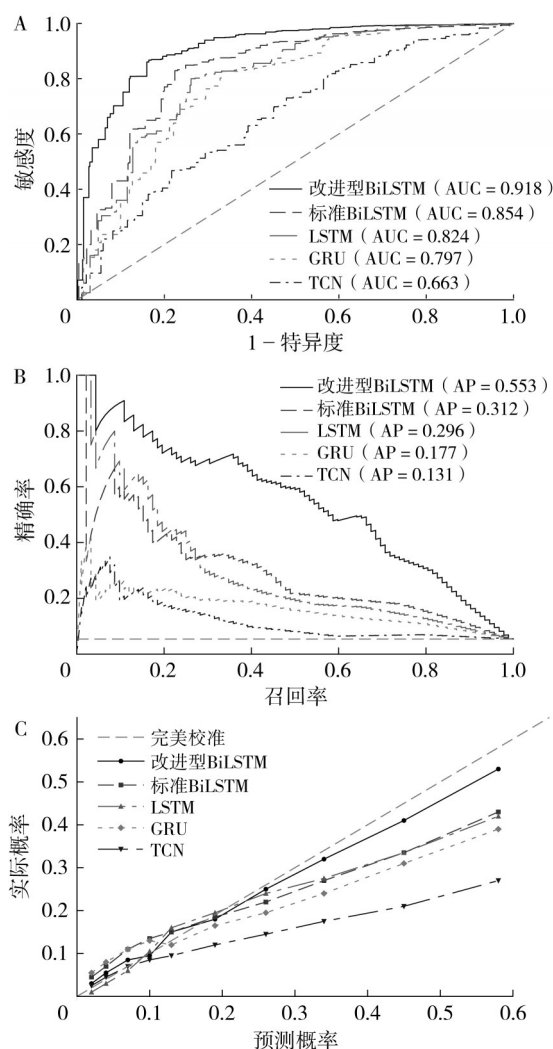


图2 测试集模型评价结果

注: A: ROC 曲线; B: PR 曲线; C: 概率校准曲线。ROC: 受试者操作特征; PR: 精确率-召回率; AP: 平均精确率; BiLSTM: 双向长短期记忆网络; AUC: 曲线下面积; LSTM: 长短期记忆网络; GRU: 门控循环单元; TCN: 时序卷积网络。

表2 测试集最优决策阈值下各预测模型的性能比较

模型类型	约登指数	准确率	敏感度	特异度	精确率	F_1 评分
改进型 BiLSTM	76.71%	95.42%	80.43%	96.28%	55.22%	65.49%
标准 BiLSTM	66.19%	94.19%	70.65%	95.53%	47.45%	56.77%
LSTM	62.55%	93.66%	67.39%	95.16%	44.29%	53.45%
GRU	57.46%	92.72%	63.04%	94.42%	39.19%	48.33%
TCN	45.23%	90.85%	52.17%	93.05%	30.00%	38.10%

注: BiLSTM: 双向长短期记忆网络; LSTM: 长短期记忆网络; GRU: 门控循环单元; TCN: 时序卷积网络。

3 讨论

急诊科多参数监护仪长期处于高频使用状态,传统基于阈值判定的维护方式难以发现质控参数在合格范围内的渐进性漂移。研究显示,本研究构建的改进型 BiLSTM 预防性维护预测模型在判别性能、少数类故障预警样本识别能力、概率校准及阈值下分类表现方面均优于参比模型,提示常规质控记录中蕴含可用于故障前识别的时序退化信息。

本研究以功能通道而非整机作为建模单元,通过滑动窗口提取跨质控周期的时序样本。该方案既可保留各功能通道独立的退化特征,又可削弱设备使用年限、维修状态变动对模型训练的干扰。BiLSTM 可同步捕捉质控序列的前后向依赖关系,较单向循环网络更适合识别短序列中的渐进性退化^[6]。引入时序注意力机制后,模型对不同质控时间步实现差异化赋权,增强对关键退化节点的感知能力,这或是本模型综合性能优于参比模型的重要内在机制^[7-8]。

从模型对比结果来看,改进型 BiLSTM 在 AUC、AP、敏感度及 F_1 评分上均优于其余参比模型,且与各参比模型的 AUC 差异均有统计学意义,提示其在整体判别能力和故障预警样本识别方面表现优异。本研究故障预警样本占比仅约 5%,处于明显类别不平衡条件下,PR 曲线较 ROC 曲线更能反映模型对少数类样本的识别能力^[9-11]。改进型 BiLSTM 的 AP 值较标准 BiLSTM 明显提高,提示时序注意力机制在稀有故障样本检测中具有明确增益作用,与既往研究结论相吻合^[12]。概率校准方面,改进型 BiLSTM 的 Hosmer-Lemeshow 检验提示校准良好,其余参比模型均存在明显校准偏差,说明其预测概率更具临床可信度,可为不同风险偏好下的维护阈值设定提供依据。与其他循环网络架构相比,LSTM 和 GRU 的表现均略于改进型 BiLSTM,进一步提示双向编码在捕捉跨质控周期时序依赖方面具有优势^[6]。相较于同领域研究^[4-5],本模型无需物联网传感数据及设备接口改造,直接利用医院现有季度质控记录,可实现通道级故障风险预警,更便于临床部署与推广。

本研究仍存在一定局限性:数据来源于单中心,样本量及品牌构成可能限制模型外部效能;故障预警样本占比较低,虽经 SMOTE 处理,但对罕见故障模式的识别能力仍有待验证;研究采用固定长度窗

口,未对比不同窗口或可变序列建模策略的效果。未来应开展多中心前瞻性研究,并结合设备使用年限、运行负荷及环境参数等信息进一步优化模型。

综上,本研究基于多参数监护仪历次质控数据,构建并验证了引入时序注意力机制的改进型 BiLSTM 预防性维护预测模型。该模型能有效识别常规质控参数中的跨周期退化趋势,在判别性能与概率校准方面均优于参比模型,可为急诊医疗设备管理从被动维修向主动预警转变提供可行的技术路径。

参考文献

- [1] 王梅,张志忠,于亚军,等. 基于模糊聚类评估技术的心电监护仪运行质量管理模型应用效果分析[J]. 中国医学装备, 2024, 21 (8): 137-141.
- [2] LI K, SU L, CHENG J, *et al.* Improving maintenance efficiency and controlling costs in healthcare institutions through advanced analytical method[J]. Sci Rep, 2025, 15(1):18377.
- [3] 李桂兰, 胡伟坡, 刘文音, 等. 医疗设备预防性维护保养实践在医疗机构的应用研究[J]. 现代医院, 2025, 25 (1): 100-103.
- [4] SHAMAYLEH A, AWAD M, FARHAT J. IoT based predictive maintenance management of medical equipment[J]. J Med Syst, 2020, 44(4):72.
- [5] RAHMAN N HABD, MOHAMAD ZAKI M H, HASIKIN K, *et al.* Predicting medical device failure: a promise to reduce healthcare facilities cost through smart healthcare management[J]. PeerJ Comput Sci, 2023, 9:e1279.
- [6] HU J, JIANG X, XU H, *et al.* Fault prediction of aircraft engine based on adaptive hybrid sampling and BiLSTM[J]. Sci Rep, 2025, 15(1):13726.
- [7] LV H, CHEN J, PAN T, *et al.* Attention mechanism in intelligent fault diagnosis of machinery: a review of technique and application[J]. Measurement, 2022, 199:111594.
- [8] DENG FENG Z, CHAOYANG T, ZHIJUN F, *et al.* Multi scale convolutional neural network combining BiLSTM and attention mechanism for bearing fault diagnosis under multiple working conditions[J]. Sci Rep, 2025, 15(1):13035.
- [9] YUAN Y, WEI J, HUANG H, *et al.* Review of resampling techniques for the treatment of imbalanced industrial data classification in equipment condition monitoring[J]. Eng Appl Artif Intell, 2023, 126:106911.
- [10] MOVAHEDI F, PADMAN R, ANTAKI J F. Limitations of receiver operating characteristic curve on imbalanced data: assist device mortality risk scores[J]. J Thorac Cardiovasc Surg, 2023, 165(4):1433-1442. e2.
- [11] OWUSU-ADJEI M, HAYFRON-ACQUAH JBEN, FRIMPONG T, *et al.* Imbalanced class distribution and performance evaluation metrics: a systematic review of prediction accuracy for determining model performance in healthcare systems[J]. PLOS Digit Health, 2023, 2(11):e0000290.
- [12] BORRÉ A, SEMAN L O, CAMPONOGARA E, *et al.* Machine fault detection using a hybrid CNN-LSTM attention-based model[J]. Sensors, 2023, 23(9):4512.