J Clin Electrocardiol, Vol.34, No.5 Oct.2025

·特约专稿·

通信作者

谢路昱，博士，副研究员。2014年获得复旦大学计算机科学与技术学士学位，2020年获得新加坡国立大学计算机科学博士学位。曾任新加坡国立癌症中心担任数据科学家。目前是合肥综合性国家科学中心数据空间研究院的副研究员，研究领域包括生物信息学、人工智能和精准医疗。研究成果发表在Bioinformatics, IEEE Transactions on Computational Biology and Bioinformatics, International Journal of Biological Macromolecules等国际高水平期刊上。

基于抄写机制的长程动态心电图数据压缩方法

周艾欣 谢路昱

【摘要】背景 长程动态心电监测在心律失常筛查与远程医疗中应用广泛，但其产生的连续高采样率信号导致数据传输与存储负担显著增加。现有心电压缩方法多针对传统心电或离线数据，难以在资源受限的可穿戴设备中实现实时、高保真压缩。**方法** 针对上述问题，本文提出了一种基于抄写机制的心电信号压缩算法TECA及其配套文件格式TEF。该方法以LZ77算法为基础框架，引入信噪比约束与滑动窗口自适应机制，构建可实时运行的容错匹配模型，从而在压缩率与波形保真度之间实现动态平衡。**结果** 在大规模Icentia 11k长程心电数据集上的验证结果表明，TECA在信噪比不低于20 dB的条件下实现了27.65的平均压缩率，百分均方根差控制在10%以内。其压缩性能显著优于小波变换、压缩感知及传统无损算法，能够有效保留心电信号的关键诊断特征。**结论** 本研究提出的TECA算法在保证高压缩率与高保真度的同时，大幅降低了心电监测设备的计算与通信负担，为长程心电仪的低功耗设计和边缘智能信号处理提供了新的技术路径，对可穿戴设备的智能化与数据高效传输具有启示意义。

【关键词】长程动态心电图；数据压缩；LZ77

［中图分类号］R541.7 R540.4 **［文献标识码］**A **［文章编号］**1005-0272(2025)05-328-06

［引用格式］周艾欣, 谢路昱. 基于抄写机制的长程动态心电图数据压缩方法[J]. 临床心电学杂志, 2025, 34(5): 328-333.

A data compression method for long-term holter ECG based on transcription mechanism *ZHOU Aixin, XIE Luyu. Institute of Dataspace, Hefei Comprehensive National Science Center, Anhui Hefei 230031, China.*

【Abstract】 Background Long-term Holter electrocardiogram (ECG) monitoring plays a crucial role in arrhythmia screening and remote healthcare. However, the continuous high-sampling-rate signals it produces lead to significant burdens in data transmission and storage. Existing ECG compression methods are mostly designed for conventional ECG or offline data and are difficult to implement in resource-constrained wearable devices to

基金项目：安徽省科技攻坚计划项目（编号：202423k09020009）

作者单位：230031 安徽 合肥，合肥国家综合性科学中心数据空间研究院

作者简介：周艾欣，硕士，主要从事医疗数据分析与人工智能相关的研究。谢路昱，博士，主要从事生物信息学与医疗人工智能方向的研究。

通信作者：谢路昱，E-mail: xieluyu@idata.ah.cn

achieve real-time, high-fidelity compression. **Methods** To address these issues, this study proposes a novel Transcription-based ECG Compression Algorithm (TECA) and its corresponding file format, the Transcription-based ECG Format (TEF). The proposed method builds upon the LZ77 framework, introducing a signal-to-noise ratio (SNR) constraint and an adaptive sliding-window mechanism to construct a fault-tolerant matching model capable of real-time operation, thereby achieving a dynamic balance between compression ratio and waveform fidelity. **Results** Experiments conducted on the large-scale Icentia 11k long-term ECG dataset demonstrate that TECA achieves an average compression ratio of 27.65 while maintaining an SNR above 20 dB and a percent root-mean-square difference (PRD) below 10%. Compared with wavelet transform, compressed sensing, and traditional lossless algorithms, TECA exhibits significantly superior compression efficiency while effectively preserving essential diagnostic ECG features. **Conclusion** The proposed TECA algorithm ensures both high compression efficiency and high fidelity, substantially reducing the computational and communication burdens of ECG monitoring devices. It provides a new technical pathway for low-power design and edge-intelligent signal processing in long-term ECG recorders, offering insights for the development of intelligent and data-efficient wearable medical devices.

【Keywords】 Holter ECG; Data compression; LZ77

1 引言

长程动态心电仪是一种能够实现连续多日心电信号采集的便携式监测设备。相较于传统动态心电仪,长程心电仪佩戴负担较小,在体积、重量及舒适度上具有明显优势,几乎不影响使用者的日常生活。因此,该设备不仅能用于临床患者的心电监测,还可作为智能穿戴设备融入健康人群的日常生活。

长程动态心电仪通常仅采集心电信号,并不具备心电数据的分析功能。所以,当这类设备用于实时检测时,通常需要将采集的数据发送至远程服务器进行存储和分析。然而,将原始的心电数据直接上传到服务器可能带来一些潜在的问题:第一,长时间的数据传输会影响心电仪的电池续航时间,从而导致监测周期缩短;第二,大量的数据传输会产生不小的数据通讯费用和存储负担,增加使用成本。为了解决这些问题,设备需在发送数据前采用压缩技术来降低通讯数据量,同时也能为分析服务器降低数据长期备份的存储负担。针对这一需求,适用于长程心电仪的数据压缩技术应当具有以下特征:

高压缩比:压缩后的数据相较于原始数据应当在字节数上大大降低。

高保真度:有损压缩技术通常能极大提升压缩比,但信号会存在一定程度的失真。因此,需要将失真控制在一定范围内,不影响后续数据分析和人工判读。

低延迟性:长程心电图的监测周期可长达一周以上,因此压缩过程应当是实时进行或至少是延迟较低的。

低计算量:由于心电仪自带的计算芯片算力有限,压缩算法的计算量应该尽量低。这也是为了节省电池的电量开销。

现有的心电信号压缩研究主要集中于常规心电图数据,方法大致可分为四类。第一类基于频域变换,如自适应傅里叶分解(Ma,2014)和小波变换(Rajoub,2002;Bera, 2019),代表性研究有 Banerjee 等(2021)在小波框架中引入可调 Q 值和混合变换结构以改善失真。但这类算法计算复杂度较高,通常依赖卷积变换运算,适合离线压缩环境,不利于实时处理;同时,频域变换对突变信号和异常心拍的保真性较差,容易造成波形细节失真。

第二类基于人工智能模型,包括自编码器(Hooshmand,2017)、卷积自编码器(Yildirim,2018;Wang,2019;Dasan,2021)以及长短期记忆网络(Hua,2022;Chauhan,2015)等。这类方法通过对大量心电数据进行预训练,能够自动提取特征、实现较高压缩比。但另一方面,这类方法通常计算量较大,难以部署在低功耗设备上,且在训练集中未出现的波形上压缩效果往往欠佳。

第三类为压缩感知方法(Picariello,2021),能够抽取出数据在频域的主要成分,并通过随机采样保证压缩性能和保真性能。但由于随机采样的限制,这类方法无法实时压缩。而且这类技术尽管保真性较好,但较其他方法压缩比更低。

第四类是通用压缩算法,如 gzip 等基于 Lempel-Ziv 系列的无损压缩技术(Horspool,1995)。

这类算法无需信号特征建模，直接对原始数据进行编码，优点是算法成熟且实现简便，但未能充分考虑心电数据的周期性和传感器噪音等特性，导致压缩率有限。

常规心电图由于数据总量较小，有数据压缩需求时常使用通用数据压缩技术即可满足需求，如 ZIP 压缩算法。但这类技术无法充分利用心电数据的特点来提高压缩率，因此在长程心电图数据上的压缩效果并不够理想。目前，关于长程心电图压缩方法的相关文献较为有限。即使在为数不多的相关论文中，大多数方法也仅针对 MIT-BIH 心律失常数据集进行了调优与测试。但该数据集采集于上世纪 70 年代，时长为 30 分钟，其信号模式特点与近年来在实践中使用的长程心电图并不相同。因此，这一领域仍缺乏基于真实场景和最新数据的探索。

对此，本研究提出了一种专门针对长程心电图信号所设计的数据压缩方法。它基于经典的 LZ77 算法进行了容错性改进以增加压缩比，同时可以设定信噪比下限。此外，该方法还能够实时编码解码，并且计算逻辑简单。基于真实长程心电图数据集的测试结果表明，在保证信噪比不低于 20 dB 的标准下，该方法的压缩率可达 20 以上。

2 方法

2.1 抄写式心电图格式

本研究首先定义一种新的心电数据压缩文件格式——抄写式心电图格式（Transcription-based ECG Format，以下简称 TEF）。该格式旨在实现长程心电信号的高效存储与解析。TEF 文件仅对传感器采集到的数字波形数据进行压缩存储，而设备的采样率、分辨率、通道数等参数可单独以元数据文件保存，从而保持格式的通用性与简洁性。

与依赖外部字典或密钥的压缩方式不同，TEF 的压缩与解压缩无需依赖任何字典或密钥，压缩文件本身即包含解压所需的全部信息，遵循 TEF 文件规范的系统均可直接进行压缩与解压，无需额外依赖。

TEF 文件由多个片段组成，片段分为原始片段和抄写片段两种类型，其类型由片段的首字节判定：当首字节作为有符号整型值小于 0 时为原始片段，大于或等于 0 时为抄写片段。原始片段用于记录那些无法有效压缩或需保留原始波形的信号。此时，首字节的绝对值表示片段长度 L，随后紧跟 L 个原始 ADC 数据信号，在解压时可直接按顺序输出这 L 个信号。若为抄写片段，则首字节以 varint 形式解析（即持续读取字节，每个字节取后 7 个比特位组成无符号整型，直至最高比特位为 1 则停止读取）并读取后续字节作为抄写偏移量 D，随后再以 varint 读取抄写长度 L。接下来再读取一个字节作为有符号整型，表示抄写时每个数值的修正值 B。解压该片段时，设当前已输出信号序列为 Z[1...N]，则调取 Z[(N-D+1)...(N-D+L)]，并对每个元素加上修正值 B，输出为解压结果。图 1 给出了一个压缩后的格式样例。

字节位置	字节内容							
1~8	-2	40	41	0	2	2	-3	-1
9~16	45	0	-5	3	10	0	-4	4
17~24	-5	-2	20	21	0	-10	8	0
25~32	0	-15	10	0				

图 1 压缩后的 TEF 文件样例（字节数组）

输出位置	输出内容							
1~8	40	41	37	38	45	50	51	47
9~16	40	45	46	42	20	21	45	50
17~24	51	47	40	45	46	42	47	40
25~32	45	46	42	20	21	45	50	51

图 2 TEF 文件解压缩后的原始文件内容（字节数组）

图 1 样例解压后的结果

2.2 基于抄写的心电图压缩算法

在 TEF 文件结构基础上，本研究提出一种基于 LZL77 改进的心电信号压缩算法——基于抄写的心电图压缩算法（Transcription-based ECG Compression Algorithm，简称 TECA）。该算法以传统 LZ77 压缩方法为核心框架，通过引入信噪比约束和滑动窗口自适应机制，实现了心电信号的高压缩率与波形保真度之间的动态平衡。

TECA 设定了三个主要可调参数用于平衡取舍压缩率、信噪比和计算量等三个性能指标，分别是：

字典区大小(BF)、信噪比下限(SNR_limit)以及扫描窗口初始大小(bsize)。字典区大小默认取 100000,用于决定算法在进行抄写匹配时的历史参考范围。更大的字典区能提高压缩率，但会导致算法的计算量变大,运行速度变慢。信噪比下限默认设为 20 dB,用于控制抄写片段的误差容忍度，以保证解压后信号的失真不影响医学判读。更高的信噪比下限可进一步减少失真,但同时会降低压缩率。扫描窗口初始大小应设置为压缩成抄写片段能够减少字节数的最短原始信号序列长度，而这取决于原始数据的采样格式,例如当信号为 16 位整型时,bsize 取 3 即可表示一个最短可压缩的心电波形单元。

TECA 算法整体框架与 LZ77 算法类似，将读入的新数据放入输入缓冲区并实时处理，压缩后的数据则输出至压缩文件。算法在执行压缩的同时同步进行解压重建,将解压后的信号放入字典区,作为后续压缩的参考基础。

在新的信号数据于输入缓冲区后,TECA 以一个初始大小为 bsize 的滑动窗口对信号进行扫描。每次迭代时，算法将窗口内的连续信号视为一个完整片段,与字典区中所有可能位置的信号序列进行匹配。对于字典区中的每一个匹配位置,TECA 通过计算对应各位置的差异均值来确定最优的抄写修正值(bias),并进一步计算局部信噪比(SNR),以评估匹配片段的相似程度。记滑动窗口的大小为 k,当前窗口信号为 V_i,信号均值为 $\bar{V}$,字典区对应参考信号为 R_j,历史信号总数为 n,局部信噪比公式如下：

$$SNR = 10 \times log_{10}[\frac{\frac{k}{n}\sum_i (V_i - \bar{V})^2}{\sum_j (V_j - R_j)^2}]$$

其中，分子部分的$\sum_i (V_i - \bar{V})^2$通过 i 枚举之前所有 n 个输入信号，估计全局信噪比分子部分的近似值，再通过系数 k/n 把全局估计值缩放到滑动窗口区域的信号个数。在实际算法中,为了避免在每次迭代中重复遍历所有历史信号，算法通过不断更新输入信号的均值与平方和，实现了信噪比分子项的快速计算。该过程可根据以下等式进行计算：

$$\sum_i (V_i - \bar{V})^2 = \sum_i (V_i^2 - 2V_i\bar{V} + \bar{V}^2) = \sum_i V_i^2 - 2\sum_i V_i\bar{V} + \sum_i \bar{V}^2 = \sum_i V_i^2 - n\bar{V}^2$$

公式分母部分的$\sum_i (V_i - \bar{V})^2$通过 j 仅枚举滑动窗口内部的信号，并计算当前窗口信号与字典区参考片段之间的误差平方和。类似的,分母部分的计算也可以通过利用与上述公式的相似的形式进行加速：为字典区中的每个位置维护当前窗口内的差值平方和与差值均值,从而在每次迭代新的修正值(记为 b)时,根据下式快速更新误差项：

$$\sum_j (V_j - (R_j + b))^2 = \sum_j ((V_j - R_j) - b)^2 = \sum_j (V_j - R_j)^2 - 2b\sum_j (V_j - R_j) + k * b^2$$

当滑动窗口移动时，差值平方和与差值均值也可以通过加减滑动窗口的头尾元素保持更新。

获得当前滑动窗口与字典区所有位置的局部信噪比之后,算法根据预设信噪比下限参数 SNR_limit 进行判断。如果字典区所有位置的局部信噪比都低于信噪比下限参数，则表明无法在满足信噪比约束的情况下,对该滑动窗口片段进行压缩。此时,算法将滑动窗口最早(即最左端,内存地址最小)的信号输入至原始片段缓存区待之后输出，并向前（即向右,内存地址更高方向)平移滑动一个信号并进入下一个迭代。如此重复,直到出现有局部信噪比高于信噪比下限，则将原始片段缓存区的信号压缩为原始片段输出。另一种情况是,如果存在一个或多个位置的局部信噪比高于阈值 SNR_limit,则表示当前滑动窗口可压缩为抄写片段。这时滑动窗口左侧不变,右侧向前扩展一个信号,进入下一个迭代,尝试压缩更长的片段。若在某次扩展过程中发现所有位置的信噪比都低于阈值，这意味着上一轮的窗口是满足信噪比约束的最大可压缩窗口，将其编码为抄写片段并输出。算法流程图如图 3。

与 LZ77 算法的“完美匹配”策略不同,TECA 通过信噪比约束允许匹配存在一定误差。这么做是因为长程单导联数据存在比普通心电图更大的随机噪声,因而能够完美匹配的片段较短,成为压缩性能的瓶颈。TECA 算法通过设置信噪比下限,允许不完美匹配的同时保证有损压缩后的数据质量。将新片段比对到字典区时，算法通过滑动窗口选取字典区内不低于信噪比下限的最长相似片段进行压缩。计算信噪比时,TECA 为字典区的每一个位置建立增量式并行扫描队列来降低计算量。

3 结果

3.1 数据集

本研究采用 Icentia 11k 数据集进行算法性能验证。该数据集是由加拿大 iCentia 公司采集并于 2022 年正式发布的大规模长程心电图数据集。它涵盖了超过 11000 名患者的连续单导联 Holter 心电监测数据,是目前规模最大、覆盖最广的长程心电图公开数据集之一。受试者佩戴 CardioSTAT 记录仪进行平均为期一周的监测,部分记录时长可达两周。所有信号均以 250 Hz 采样率、16 位分辨率采集,保证了高时间精度与高动态范围。数据集中的心拍标注由 20 名经过专

J Clin Electrocardiol, Vol.34, No.5 Oct.2025

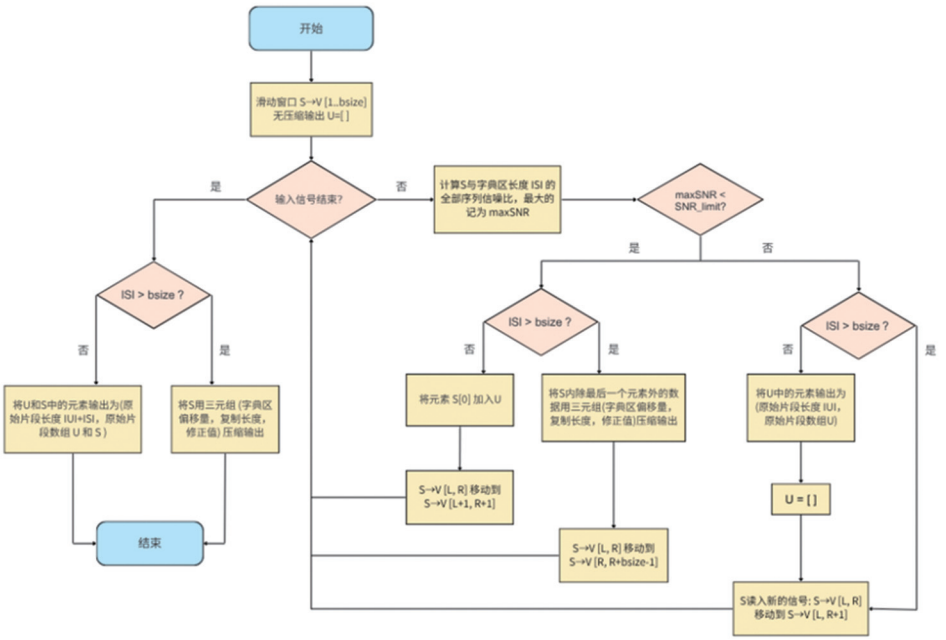

图 3 算法流程图

业认证的技术人员完成，标注内容包括心跳类型(如正常搏动、房性早搏 PAC、室性早搏 PVC)和心律类型(如窦性心律、心房颤动 AF、心房扑动 AFL)。

3.2 评估指标

为全面评估算法性能，本研究采用压缩率(Compression Ratio，CR)、信噪比 (Signal-to-Noise Ratio，SNR)、百分均方根差(Percent Root-mean-square Difference，PRD)等三个指标作为主要评价标准。

3.2.1 压缩率

压缩率是原始数据大小与压缩后数据大小的比值，数值越大表示压缩效率越高，被广泛接受为压缩效率的标准。其具体公式如下：

$$CR = \frac{压缩前数据大小}{压缩后数据大小}$$

3.2.2 信噪比

信噪比是原始信号中心化值与压缩重建后信号误差的比值，用于比较传感器噪声量级与期望的实际信号量级。SNR 越高，说明信号保留得越好。若用 V 表示输入信号，R 表示解压后信号，则 SNR 的公式可表示为：

$$SNR = 10 \times log_{10}[\frac{\sum_i (V_i - \bar{V})^2}{\sum_i (V_i - R_i)^2}]$$

3.2.3 百分均方根差

百分均方根差是解压后信号误差均方根与原始信号均方根的百分比，用于衡量解压后信号偏离原始信号的程度。PRD 越小，说明解压后信号的质量越好。其具体公式如下：

$$PRD = 100 \times \sqrt{\frac{\sum_i (V_i - R_i)^2}{\sum_i {V_i}^2}]}$$

3.3 测试结果

为验证 TECA 算法的有效性，本研究将其与多种典型心电压缩方法进行对比，包括小波变换、压缩感知及传统无损压缩算法 gzip(DEFLATE)。所有算法均在相同采样率和分辨率条件下运行，以保证结果可比性。测试结果如表 1 所示。

表 1 各压缩方法性能对比表

压缩方法	CR	SNR	PRD
TECA	27.65	20.00 dB	9.97 %
小波变换	8.2	10.58 dB	33.6 %
压缩感知	3.79	14.6 dB	18.6 %
gzip(DEFLATE)	4.5	无损压缩	无损压缩

结果表明，TECA 在保证信噪比达到 20 dB 的条件下，实现了 27.65 的压缩率，显著优于小波变换与压缩感知等主流方法，压缩效率提升约 3~7 倍。同时，其 PRD 控制在 10% 以内，能够有效保留心电信号的波形细节和关键诊断特征。相较之下，小波变换虽能在低频区保持一定保真度，但在突变波形与异常节律段存在较大失真；压缩感知和 gzip 压缩率较低，难以满足长程监测设备的能耗与存储要求。

4 讨论

本研究针对长程动态心电监测中存在的数据传输与存储压力问题，提出了一种基于抄写机制的心电信号压缩方法(TECA)及其配套文件格式(TEF)。该方法以 LZ77 为基础框架,通过引入信噪比约束与滑动窗口策略,在显著降低计算与通信成本的同时,保持了波形的重要特征，为后续异常检测的灵敏度与可靠性提供保障。

从算法结构上看,TECA 的创新点主要体现在三个方面:其一,利用“抄写片段”机制重构局部相似波形,使压缩过程可直接在设备端实时执行;其二,通过信噪比约束策略，将传统意义上的字典匹配问题转化为基于信号保真度的动态匹配问题，提升了压缩率和鲁棒性;其三,引入滑动窗口快速更新与局部统计量缓存机制,有效降低了计算负担,使算法能够在资源受限的嵌入式芯片上实现。

实验结果表明,TECA 在 Icentia 11k 大规模长程心电数据集上取得了显著性能优势。在保证信噪比不低于 20 dB 的条件下，平均压缩率达到 27.65,PRD控制在 10% 以内,明显优于小波变换和压缩感知等主流方法,且压缩质量满足临床判读的要求。该结果表明,TECA 能够在极大减少通信与存储负担的同时,保留关键的心电波形特征,对长程监测设备的能效设计和边缘计算策略具有重要意义。该方法为长程动态心电图数据提供了新的压缩解决方案,延长设备续航时间并减轻服务器负担。在可穿戴健康设备领域,TECA 提供了兼顾实时性与低功耗的轻量化压缩思路，为连续生理监测与边缘健康分析奠定基础。在应用层面,其无字典、可独立解压的特性使其具备良好的平台兼容性，便于与云端分析系统或移动端终端结合,实现实时的信号压缩与解析。

尽管本研究提出的 TECA 算法在压缩率与信号保真度方面均取得了良好效果，但仍存在若干有待改进的方面。例如,当前研究的主要目标是实现高效压缩与高保真重建，尚未在压缩过程中引入临床语义层面的信息。未来研究可围绕在线异常信号检测与实时诊断结果输出展开探索。通过结合压缩算法与心律分类模型，可在压缩过程中同步完成异常节律的识别与标注，实现数据压缩与初步诊断的一体化设计。这不仅有助于减少原始数据传输量,还能显著提升可穿戴心电设备在远程医疗与即时健康监测中的智能化水平。

参考文献

[1] MA J L, ZHANG T T, DONG M C. A novel ECG data compression method using adaptive fourier decomposition with security guarantee in e-health applications [J]. IEEE J Biomed Health Inform, 2014, 19(3): 986-994.

[2] RAJOUB B A. An efficient coding algorithm for the compression of ECG signals using the wavelet transform[J]. IEEE Trans Biomed Eng, 2002, 49(4): 355-362.

[3] BERA P, GUPTA R, SAHA J. Preserving abnormal beat morphology in long-term ECG recording: an efficient hybrid compression approach [J]. IEEE Trans Instrum Meas, 2019, 69(5): 2084-2092.

[4] BANERJEE S, SINGH G K. Quality guaranteed ECG signal compression using tunable-Q wavelet transform and Möbius transform-based AFD[J]. IEEE Trans Instrum Meas, 2021, 70: 1-11.

[5] HOOSHMAND M, ZORDAN D, DEL TESTA D, et al. Boosting the battery life of wearables for health monitoring through the compression of biosignals [J]. IEEE Internet of Things J, 2017, 4(5): 1647-1662.

[6] YILDIRIM O, TAN R S, ACHARYA U R. An efficient compression of ECG signals using deep convolutional autoencoders [J]. Cogn Syst Res, 2018, 52: 198-211.

[7] WANG F, MA Q M, LIU W H, et al. A novel ECG signal compression method using spindle convolutional auto-encoder[J]. Comput Methods Programs Biomed, 2019, 175: 139-150.

[8] DASAN E, PANNEERSELVAM I. A novel dimensionality reduction approach for ECG signal via convolutional denoising autoencoder with LSTM [J]. Biomed Signal Process Control, 2021, 63: 102225.

[9] HUA J, RAO J, PENG Y Q, et al. Deep compressive sensing on ECG signals with modified inception block and LSTM[J]. Entropy, 2022, 24(8): 1024.

[10] CHAUHAN S, VIG L. Anomaly detection in ECG time signals via deep long short-term memory networks [C]//2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA). IEEE, 2015: 1-7.

[11] PICARIELLO F, IADAROLA G, BALESTRIERI E, et al. A novel compressive sampling method for ECG wearable measurement systems[J]. Measurement, 2021, 167: 108259.

[12] HORSPOOL R N, WINDELS W J. ECG compression using Ziv-Lempel techniques[J]. Comput Biomed Res, 1995, 28(1): 67-86.

（收稿日期:2025-09-20）