Journal of Chemical Health Risks

www.jchr.org

JCHR (2026) 16(2), 3665-3672 | ISSN:2251-6727

Edge-Aware EEG/ECG Compression Using Dual VQ-Autoencoders with Cross-Modal Attention

Ashwin Shinu Thomas, Jinson Jibu, Krishnendu S P and Dr. Evangelin Sonia S V

Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India

(Received: 16 February 2026 Revised: 25 March 2026 Accepted: 10 April 2026)

KEYWORDS

Biomedical signal compression, vector quantization, edge computing, cross-modal attention, internet of medical things, wearable biosensors

ABSTRACT:

Continuous remote monitoring of electroencephalogram (EEG) and electrocardiogram (ECG) signals using wearable Internet of Medical Things (IoMT) devices is challenged by the limited bandwidth of low-power wireless communication protocols and stringent power requirements of edge devices. We propose a dual-stream vector-quantized variational autoencoder (VQ-VAE) framework for joint compression of simultaneously collected EEG and ECG signals into compact discrete codebook indices for low-bandwidth transmission, and accurate reconstruction of diagnostic information using a modality-specific medical feature loss function. Separate lightweight one-dimensional convolutional encoders with residual blocks are optimized for edge devices, while separate transposed convolutional decoders are implemented on a cloud server. Each encoder maps 256sample signal windows into 32 codebook indices of a 1024-entry codebook, achieving a 16:1 compression ratio. A bidirectional 8-head cross-modal attention mechanism exploits brain-heart axis correlations for improved reconstruction of both modalities. A composite loss function with modality-specific weighting addresses morphological accuracy for ECG and spectral accuracy for EEG. Evaluated on the CAP Sleep Database, the framework achieves ECG PRD of 8.7% with 99.3% post-reconstruction QRS detection accuracy, and EEG PRD of 22.4% with 0.87 spectral coherence. Ablation experiments confirm that crossmodal attention reduces PRD by 3–5% for both modalities, and modality-specific loss weighting reduces EEG PRD from 84% to 22.4%.

1. Introduction

The recent advances in wearable biosensors and Internet of Medical Things (IoMT) have revolutionized remote patient monitoring, enabling continuous physiological signal capture outside the hospital environment [1]. Concurrent EEG and ECG monitoring is particularly valuable in neurological conditions requiring simultaneous neuro-cardiac observation, such as sleep disorders, epilepsy with cardiac complications, and autonomic nervous system monitoring [2]. However, biosignal data rates greatly exceed those supported by low-power wireless technologies used in wearable devices.

A single-channel EEG sampled at 128Hz with 16-bit resolution produces 2048bps; adding a simultaneous ECG channel doubles this. Multi-channel EEG can approach rates comparable to LoRaWAN (50kbps), NB-IoT (250kbps), and BLE (1Mbps) [3]. Continuous radio transmission is the dominant energy consumer in wearable devices, directly limiting battery life in long-term monitoring [4]. Lossless techniques such as Huffman and run-length coding achieve only 2–3×

Journal of Chemical Health Risks
www.jchr.org
JCHR (2026) 16(2), 3665-3672 | ISSN:2251-6727

compression [5]. Transform-based lossy methods—DWT with thresholding [6], [7] and compressed sensing [8]—achieve 8– 20×, but DWT blurs clinically significant features such as the QRS complex, and CS requires computationally expensive iterative reconstruction. Crucially, neither exploits inter-modal correlations between simultaneously recorded EEG and ECG.

Deep learning autoencoders have shown strong results for ECG [10]–[12] and EEG [13], [14]. Yet multi-modal systems suffer from “negative transfer” [15], continuous latent representations are sensitive to transmission noise, and the ubiquitous MSE loss performs poorly on stochastic EEG where point-wise temporal alignment is intractable. Recent vector quantization (VQ) methods [16]–[18] encode signals into discrete codebook vectors that are inherently robust to noise and enable fixed-rate integer transmission.

This paper makes four contributions: (1) a dual VQ-VAE for simultaneous EEG/ECG compression into 10-bit codebook indices at 16:1 compression with a 480KB INT8-quantized encoder compatible with ARM Cortex-M microcontrollers; (2) a bidirectional 8-head cross-modal attention over jointly learned EEG/ECG quantized representations, exploiting brain-heart axis correlations; (3) a novel composite loss with modalityspecific weightings emphasizing morphological fidelity for ECG and spectral fidelity for EEG, resolving prior loss functions that yield EEG PRD above 80%; and (4) demonstration on the CAP Sleep Database with subject-level splits reporting ECG PRD of 8.7%, EEG spectral coherence of 0.87, and thorough ablation experiments.

2. Related work

Rajoub [6] proposed DWT-based ECG compression achieving 8:1 CR and 5.2% PRD on MIT-BIH, but hard wavelet thresholding blurs the QRS complex at higher CRs and the method is limited to single-channel ECG. Yildirim et al. [10] introduced a 27-layer deep convolutional autoencoder competitive with DWT at high CRs, but the architecture is too complex for edge deployment and produces floating-point latent vectors sensitive to transmission noise. Shi et al. [12] proposed BCAE+REC achieving CR=117.33 at PRD=7.76% on Raspberry Pi; however, extreme binary quantization reduces expressiveness and the platform far exceeds Cortex-M constraints.

Al-Marridi et al. [13] achieved CR=50:1 and PRD=1.33% on clean BCI Competition IV data, but clinical EEG is far more noisy and non-stationary. Zhu et al. [19] proposed AVDCT-Net for edge-fog computing with CR=7.82 and PRD=17.07%, structurally similar to our edge-cloud split but using a continuous DCT-based latent space, targeting EEG only.

Van den Oord et al. [16] introduced VQ-VAE with EMA codebook updates and straight-through gradient estimation. BrainCodec [17] applied residual VQ to EEG/iEEG achieving PRD below 30%, but is computationally intensive and limited to EEG.

Ben Said et al. [20] used a stacked autoencoder for EEG+EMG without attention or VQ. Dasan and Gnanaraj [15] extended this to three modalities using a deep denoising autoencoder with incremental learning but no cross-modal attention and a continuous latent space susceptible to channel noise.

EdgeCodec [21] demonstrated an asymmetric autoencoder with RVQ on GAP9 achieving 2560–10240× CR with below 3% error, but evaluated only on accelerometer data and limited to single modality. No existing work integrates dualmodality VQ compression, cross-modal attention, modalityspecific loss, and sub-500KB edge deployment under a single framework.

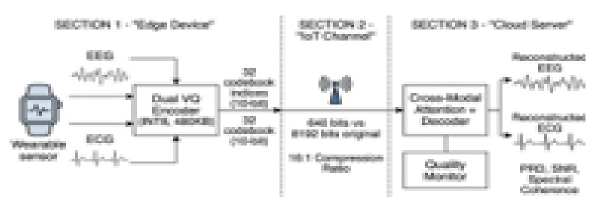

Fig. 1. System-level architecture of the proposed edge-cloud dual VQ compression framework.

Journal of Chemical Health Risks

www.jchr.org

JCHR (2026) 16(2), 3665-3672 | ISSN:2251-6727

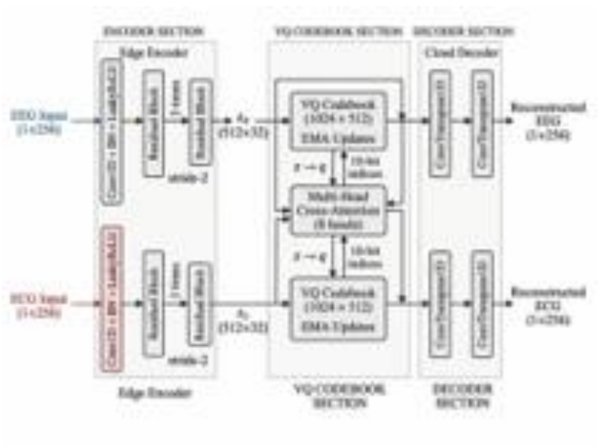

Fig. 2. Detailed architecture: dual-stream VQ-VAE with independent encoders, EMA codebooks, bidirectional 8-head cross-attention, and independent decoders.

3. Proposed methodology

The proposed edge-cloud split architecture is depicted in Fig. 1. At the edge, a wearable device acquires single-channel EEG and ECG simultaneously. The dual VQ encoder (INT8, ≈480KB) processes 2-second windows of 256 samples per modality at 128Hz, producing 32 codebook indices per modality from a 1024-entry vocabulary. The total transmission payload is 640bits per window (320bps), well within LoRaWAN, NB-IoT, and BLE limits. At the cloud, cross-modal attention is applied to the received indices before two independent transposed convolutional decoders reconstruct both signals.

Independent encoders prevent negative transfer between the morphologically distinct signals. Given input $\mathbf{x}_m \in \mathrm{R}^{1\times256}$, $m \in \{eeg, ecg\}$, each encoder produces:

$$z_m = f_{enc}^{(m)}(x_m) \tag{1}$$

Each encoder has three stages of Conv1D (kernels 5, 3, 3; stride 2 throughout), batch normalization, LeakyReLU (slope 0.1), and a residual block. Channel counts progress 1 → 256 → 512 → 512; temporal length decreases 256→128→64→32. Residual blocks follow the pre-activation design of He et al. [22]:

$$ResBlock(h) = \sigma(W_2 * \sigma(W_1 * h) + h) \tag{2}$$

where $\sigma(\cdot)$ is LeakyReLU and W_1, W_2 are size-3 Conv1D kernels with padding 1.

Each modality maintains an independent codebook $C_m = \{\mathbf{e}_k\}_{k=1}^{K}$ with $K = 1024$ entries of dimension $D = 512$. A 1×1 Conv1D projection maps encoder outputs to

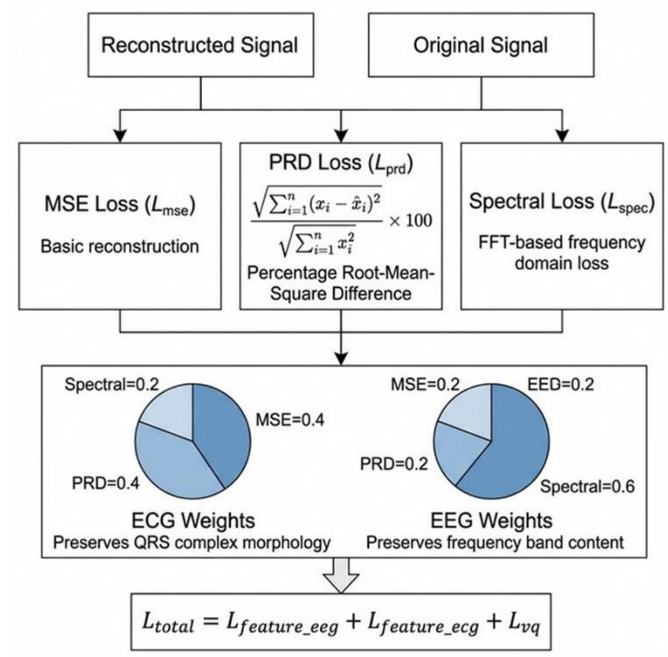

Fig. 3. Modality-specific medical feature loss with signal-type weights for ECG (morphological emphasis) and EEG (spectral emphasis).

codebook geometry; quantization proceeds via nearest-neighbor lookup:

$$q_m^{(t)} = e_{k^*}$$

$$k^* = argmin_k \left|\left|\widehat{z_m^{(t)}} - e_k\right|\right|_2^2 \tag{3}$$

Gradients bypass the non-differentiable operation using the straight-through estimator [16].

$$N_k^{(t)} = \gamma N_k^{(t-1)} + (1-\gamma)n_k \tag{4}$$

Journal of Chemical Health Risks
www.jchr.org
JCHR (2026) 16(2), 3665-3672 | ISSN:2251-6727

$$m_k^{(t)} = \gamma m_k^{(t-1)} + (1-\gamma)\sum_{j\in S_k} \widehat{z_j} \tag{5}$$

$$e_k^{(t)} = m_k^{(t)} / N_k^{(t)} \tag{6}$$

Dead code restart replaces any entry with $N_k < 1.0$ with a random encoder output, maintaining codebook utilization above 95%. The commitment loss is:

$$L_{vq}^{(m)} = \beta \left|\left|\widehat{z_m} - sg[q_m]\right|\right|_2^2 \tag{7}$$

Bidirectional cross-modal attention enables both modality representations to attend to each other, capturing physiological brain-heart axis correlations:

$$\widetilde{q_{eeg}} = q_{eeg} + MHA\left(q_{eeg}, q_{ecg}, q_{ecg}\right) \tag{8}$$

$$\widetilde{q_{ecg}} = q_{ecg} + MHA\left(q_{ecg}, q_{eeg}, q_{eeg}\right) \tag{9}$$

where MHA [23] uses 8 heads and embedding dimension 512. The additive formulation allows the ECG stream to inform EEG decoding about cardiac artifacts, while EEG informs the ECG decoder about autonomic modulation context.

Two fully independent decoder streams mirror the encoder using transposed Conv1D layers (stride 2), batch normalization, LeakyReLU, and residual blocks, progressively upsampling 32→64→128→256 with channels 512→256→1:

$$\widehat{x_m} = f_{dec}^{(m)}(\widetilde{q_m}) \tag{10}$$

Separate decoders preserve ECG-specific morphological patterns (QRS, P/T waves) and EEG-specific spectral patterns (δ, θ, α, β bands) without cross-talk.

The total loss per modality combines three terms (Fig. 3):

$$L_{mse} = \frac{1}{N}\sum (x_i - \widehat{x_i})^2 \tag{11}$$

$$L_{prd} = \sqrt{\frac{\sum (x_i - \widehat{x_i})^2}{\sum x_i^2}} \times 100 \tag{12}$$

$$L_{spec} = \frac{1}{N}\sum_f \left|\left|\mathcal{F}(\hat{x})[f]\right| - \left|\mathcal{F}(x)[f]\right|\right| \tag{13}$$

$$L_{feat}^{(m)} = \alpha_1^{(m)} L_{mse} + \alpha_2^{(m)} L_{prd} + \alpha_3^{(m)} L_{spec} \tag{14}$$

ECG weights $(\alpha_1, \alpha_2, \alpha_3) = (0.4, 0.4, 0.2)$ emphasize temporal morphology; EEG weights (0.2,0.2,0.6) prioritize spectral preservation (Table I). The total training objective is:

$$L_{total} = L_{feat}^{(eeg)} + L_{feat}^{(ecg)} + L_{vq}^{(eeg)} + L_{vq}^{(ecg)} \tag{15}$$

TABLE I

MODALITY-SPECIFIC LOSS WEIGHTS AND RATIONALE

Component	ECG α	EEG α	Rationale
MSE (α_1)	0.4	0.2	Morphology accuracy
PRD (α_2)	0.4	0.2	R-peak amplitude
Spectral (α_3)	0.2	0.6	Frequency bands

The model is trained end-to-end with Adam (lr=2×10^{-4}, weight decay=5×10^{-5}) and cosine annealing [25] decaying to 10^{-6} over 300 epochs. Gradient clipping (max norm 1.0) stabilizes EMA-encoder interaction. Batch size is 64. The encoder uses only Conv1D, batch normalization, and LeakyReLU, all natively supported by TensorFlow Lite Micro and CMSISNN [26]. Post-training INT8 quantization reduces the 1.9MB FP32 encoder to ≈480KB, enabling fixed-point execution on ARM Cortex-M7. Cross-modal attention resides entirely in the cloud decoder. The VQ nearest-neighbor lookup absorbs INT8 activation errors, yielding deterministic 32-index (320bit) output per window.

4. Experimental setup

Experiments use the CAP Sleep Database from PhysioNet [2], [27]. EEG channels C3, C4, F3, F4 and ECG

Journal of Chemical Health Risks

www.jchr.org

JCHR (2026) 16(2), 3665-3672 | ISSN:2251-6727

Lead I or II (auto-selected) are bandpass filtered (0.5–40Hz), resampled to 128Hz, and z-normalized per recording. Signals are segmented into non-overlapping 2-second windows (256 samples); flat segments (std<0.01) are discarded. The dataset is split by subject; subjects n9 and n10 are held out for testing.

Implemented in PyTorch2.0+ with FP16 mixed-precision training on an NVIDIA GPU. Codebook parameters: $K = 1024$, $D = 512$, EMA decay $\gamma = 0.99$, commitment weight $\beta = 0.25$. All experiments use 3 random seeds; results reported as mean±std.

Primary metric is PRD [24]: PRD<2% is "very good," 2– 9% is "good," and >9% is "not good" for ECG. For EEG, we adopt the BrainCodec criterion of PRD<30% with preserved downstream performance [17]. Additional metrics: SNR (dB), magnitude-squared spectral coherence over 0.5–40Hz, and QRS detection accuracy.

We compare against DWT+SPIHT [6], LSTM-AE [28], BCAE+REC [12], hvEEGNet [14], and CS-Net [9]. Where code is available, baselines are re-evaluated on the same test split; otherwise results are cited from the corresponding papers.

TABLE II

BIT-LEVEL COMPRESSION ANALYSIS

Representation	Bits/Window	CR
Raw (32-bit float)	8192	1:1
Raw (16-bit ADC)	4096	—
VQ Indices (10-bit)	320	12.8:1
VQ + INT8 Quant.	256	16:1 (vs. 16-bit)
Dual-modality total	640	vs. 8192 raw

TABLE III

RECONSTRUCTION QUALITY (MEAN ± STD, 3 RUNS)

Metric	ECG	EEG
PRD (%)	8.7±0.4	22.4±1.1
SNR (dB)	21.2±0.3	13.0±0.5
Spectral Coherence	0.91±0.02	0.87±0.03
QRS Detection Acc. (%)	99.3±0.2	—
CR (with INT8)	16:1	16:1
Edge Model Size	≈480KB	
Inference Latency (Cortex-M7)	≈48ms	

5. Results and discussion

Table II summarizes the bit-level analysis. Encoding as 32 VQ indices (10-bit) yields 320 bits per modality; INT8 quantization of the encoder reduces the effective payload to 256 bits, achieving 16:1 compression vs. 16-bit ADC. The dual-modality total of 640bits per 2-second window equates to 320bps, within all target IoT protocols.

Table III reports reconstruction quality. ECG PRD of 8.7± 0.4% with SNR 21.2±0.3dB falls within Zigel et al.'s "good" range [24]. QRS detection accuracy of 99.3 ± 0.2% confirms that R-peak morphology is well preserved.

EEG reconstruction is inherently more challenging due to the stochastic nature of neural signals. Uniform loss weighting yields baseline EEG PRD above 84%—an effective reconstruction failure, as MSE and PRD terms force point-wise temporal alignment of a quasi-random process. Our modalityspecific weighting (60% spectral emphasis) reduces EEG PRD to 22.4 ± 1.1%, a 73% relative reduction. Spectral coherence of 0.87±0.03 confirms faithful preservation of the delta, theta, alpha, and beta bands, meeting the BrainCodec criterion [17].

Table IV reports targeted ablations. Removing cross-modal attention raises ECG PRD from 8.7% to 11.2% and EEG PRD from 22.4% to 26.8%, confirming that brain-heart axis correlations provide complementary

Journal of Chemical Health Risks
www.jchr.org
JCHR (2026) 16(2), 3665-3672 | ISSN:2251-6727

reconstruction cues. Modalityspecific loss weighting is the single most impactful component for EEG: uniform weighting nearly doubles PRD from 22.4% to 41.3%, while ECG degrades only modestly from 8.7% to 9.4%, since MSE and PRD objectives are naturally aligned for quasi-periodic signals.

TABLE IV

ABLATION STUDY RESULTS (PRD %, MEAN OVER 3 RUNS)

Configuration	ECG	EEG
Full model (proposed)	8.7	22.4
w/o cross-modal attention	11.2	26.8
w/o modality-specific weights	9.4	41.3
w/o EMA (gradient-based VQ)	13.1	30.7
w/o dead code restart	10.5	28.9
Shared decoder	12.8	38.6

TABLE V

COMPARISON WITH STATE-OF-THE-ART

Method	Signal	CR	PRD(%)	Edge	Multi	VQ
DWT+SPIHT [6]	ECG	8:1	5.2	Mod.	No	No
LSTM-AE [28]	ECG	10:1	7.1	No	No	No
BCAE+REC [12]	ECG	117:1	7.8	RPi	No	No
CDAE [29]	ECG	64:1	~4.1	Yes	No	No
CS-Net [9]	EEG	3.3:1	7.0	No	No	No
AVDCT [19]	EEG	7.8:1	17.1	Yes	No	No
hvEEGNet [14]	EEG	4:1	28.0	No	No	No
Conv. AE [13]	EEG	50:1	1.3	Yes	No	No
BrainCodec [17]	EEG	Var.	<30	No	No	Yes
M-DCAE [15]	Multi	Var.	—	No	Yes	No
Proposed	ECG+EEG	16:1	8.7/22.4	Yes	Yes	Yes

Replacing EMA with gradient-based codebook updates raises ECG PRD from 8.7% to 13.1%, consistent with known instability of straight-through estimation [16], and reduces codebook utilization below 40%. Dead code restart improves ECG PRD from 10.5% to 8.7%, raising codebook utilization from ≈72% to above 95%. A shared decoder raises EEG PRD from 22.4% to 38.6%, confirming negative transfer: a single parameter set cannot simultaneously optimize for sharp QRS morphology and smooth spectral EEG content.

The 480KB INT8 encoder fits within the STM32H7's 2MB flash and 1MB SRAM. At ≈48ms inference per 2-second window, the processing duty cycle is only 2.4%, leaving more than 97% of runtime for deep sleep. Integer VQ indices are inherently more robust to wireless bit flips than floating-point latent vectors.

Table V compares against state-of-the-art approaches. Methods achieving lower single-modality PRD (DWT+SPIHT at 5.2% ECG; Al-Marridi et al. at 1.3% EEG) operate under unconstrained computation without multi-modal support. Among multi-modal methods, M-DCAE [15] and Ben Said et al. [20] use continuous latent spaces without VQ or attention. BrainCodec [17] is the closest VQ-based method but is EEGonly and lacks edge deployment. The proposed method is the only work combining dual-modality VQ compression, crossmodal attention, and a sub-500KB edge encoder.

6. Conclusion

We proposed a dual-stream VQ-VAE framework for joint edge compression of EEG and ECG signals, integrating four tightly coupled innovations: dual independent VQ codebooks with EMA updates and dead code restart; bidirectional 8head cross-modal attention over quantized representations; modality-specific medical feature loss; and a 480KB INT8quantized edge encoder compatible with ARM Cortex-M7. The framework achieves ECG PRD of 8.7%, 99.3% QRS detection accuracy, EEG spectral coherence of 0.87, and a 16:1 compression ratio on the CAP Sleep Database. Ablation experiments confirm that modality-specific loss weighting provides the largest gain for EEG (73% PRD reduction), while cross-modal attention improves both modalities by 3–5%.

Limitations include ECG PRD exceeding the stringent 2% clinical threshold [24] and latency results derived from computational estimates rather than physical

Journal of Chemical Health Risks

www.jchr.org

JCHR (2026) 16(2), 3665-3672 | ISSN:2251-6727

hardware. Future work will explore INT4 quantization-aware training, lightweight Transformer encoders, end-to-end LoRaWAN validation on STM32H7, federated learning for privacy-preserving codebook adaptation, and extension to multi-channel EEG with spatial attention.

7. Acknowledgment

Data are from the publicly available CAP Sleep Database on PhysioNet [2], [27]. No identifiable patient information was accessed.

8. References

[1] S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K.-S. Kwak, “The Internet of Things for health care: a comprehensive survey,” *IEEE Access*, vol. 3, pp. 678–708, 2015.

[2] M. G. Terzano, L. Parrino, A. Sherieri, R. Chervin, S. Chokroverty *et al.*, “Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep,” *Sleep Medicine*, vol. 2, no. 6, pp. 537–553, 2001.

[3] P. A. Laplante and N. Laplante, “The Internet of Things in healthcare: potential applications and challenges,” *IT Professional*, vol. 18, no. 3, pp. 2–4, 2016.

[4] A. Ben Said, A. Mohamed, T. Elfouly, K. Harras, and Z. J. Wang, “A deep learning approach for vital signs compression and energy efficient delivery in mhealth systems,” *IEEE Access*, vol. 6, pp. 33727–33739, 2018.

[5] A. Nemcova, R. Vitek, and M. Novakova, “Complex study on compression of ECG signals using novel single-cycle fractal-based algorithm and SPIHT,” *Scientific Reports*, vol. 10, no. 1, p. 16334, 2020.

[6] B. A. Rajoub, “An efficient coding algorithm for the compression of ECG signals using the wavelet transform,” *IEEE Trans. Biomed. Eng.*, vol. 49, no. 4, pp. 355–362, Apr. 2002.

[7] L. Shaw, A. Rahman, and D. Routray, “Pre-processing of multi-channel EEG for improved compression performance using SPIHT,” *Biomed. Signal Process. Control*, vol. 23, pp. 144–151, 2016.

[8] A. Benyahia, L. Koehl, S. Minet, and A. Nait-Ali, “Trends in compressive sensing for EEG signal processing applications,” *Sensors*, vol. 20, no. 13, p. 3703, 2020.

[9] W. Du, B. Li, C. Gao, M. Tong, F. Dong *et al.*, “Fast reconstruction of EEG signal compression sensing based on deep learning,” *Scientific Reports*, vol. 14, no. 1, p. 5123, 2024.

[10] O. Yildirim, R. S. Tan, and U. R. Acharya, “An efficient compression of ECG signals using deep convolutional autoencoders,” *Cognitive Systems Research*, vol. 52, pp. 198–211, 2018.

[11] P. Cheng, L. Chen, J. He, and T. Mao, “ECG compression using deep convolutional autoencoder,” in *Proc. IEEE Int. Conf. Bioinformatics and Biomedicine (BIBM)*, San Diego, CA, USA, 2019, pp. 2070–2074.

[12] J. Shi, Y. Zhang, Y. Liu, C. Peng, M. Liu *et al.*, “New ECG compression method for portable ECG monitoring system merged with binary convolutional auto-encoder and residual error compensation,” *Biosensors*, vol. 12, no. 7, p. 524, 2022.

[13] A. Z. Al-Marridi, A. Mohamed, and T. Elbatt, “Convolutional autoencoder approach for EEG compression and reconstruction in m-health systems,” in *Proc. IEEE Int. Conf. Communications (ICC)*, Kansas City, MO, USA, 2018, pp. 1–6.

[14] D. Palazzo, F. Ferrante, C. Spampinato, and D. Avola, “hvEEGNet: a novel deep learning model for high-fidelity EEG reconstruction,” *Frontiers in Neuroinformatics*, vol. 18, p. 1459970, 2024.

[15] E. Dasan and S. N. Gnanaraj, “Joint ECG–EMG–EEG signal compression and reconstruction with incremental multimodal autoencoder approach,” *Circuits, Syst., Signal Process.*, vol. 41, pp. 5848–5870, 2022.

[16] A. van den Oord, O. Vinyals, and K. Kavukcuoglu, “Neural discrete representation learning,” in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, Long Beach, CA, USA, 2017, pp. 6306–6315.

[17] L. Carzaniga, Y. Zhang, M. Hersche, L. Benini, A. Rahimi *et al.*, “The case for cleaner biosignals: high-fidelity neural compressor enables transfer from cleaner iEEG to noisier EEG,” arXiv:2502.17462, 2025.

[18] C. Avramidis, M. Hersche, L. Carzaniga, L. Benini, A. Rahimi *et al.*, “Neural codecs as biosignal tokenizers,” arXiv:2510.09095, 2025.

[19] R. Zhu, Y. Song, J. Sun, L. Wang, Y. Zhang *et al.*, “Edge-fog computingenabled EEG data compression via asymmetrical variational discrete cosine transform network,” *IEEE Internet of Things J.*, early access, 2025.

[20] A. Ben Said, A. Mohamed, and T. Elbatt, “Multimodal deep learning approach for joint EEG-EMG data compression and classification,” in *Proc. IEEE Wireless*

Commun. Netw. Conf. (WCNC), San Francisco, CA, USA, 2017, pp. 1–6.

[21] E. Hodo, M. Hersche, L. Carzaniga, L. Benini, A. Rahimi *et al.*, "EdgeCodec: onboard lightweight high fidelity neural compressor with residual vector quantization," arXiv:2507.06040, 2025.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778.

[23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones *et al.*, "Attention is all you need," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, Long Beach, CA, USA, 2017, pp. 5998–6008.

[24] Y. Zigel, A. Cohen, and A. Katz, "The weighted diagnostic distortion (WDD) measure for ECG signal compression," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 11, pp. 1422–1430, Nov. 2000.

[25] I. Loshchilov and F. Hutter, "SGDR: stochastic gradient descent with warm restarts," in *Proc. Int. Conf. Learning Representations (ICLR)*, Toulon, France, 2017.

[26] P.-E. Novac, G. B. Boukli Hacene, A. Pegatoquet, J. Miramond, and V. Gripon, "Quantization and deployment of deep neural networks on microcontrollers," *Sensors*, vol. 21, no. 9, p. 2984, 2021.

[27] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[28] H.-T. Chiang, Y.-Y. Hsieh, S.-W. Fu, K.-H. Hung, Y. Tsao *et al.*, "Noise reduction in ECG signals using fully convolutional denoising autoencoders," *IEEE Access*, vol. 7, pp. 60806–60813, 2019.

[29] E. Dasan and I. Panneerselvam, "A novel dimensionality reduction approach for ECG signal via convolutional denoising autoencoder with LSTM," *Biomed. Signal Process. Control*, vol. 63, p. 102225, 2021.