International Journal of Data Informatics and Intelligent Computing (IJDIIC)
Vol. 5, No. 2, June 2026, pp. 62~77
DOI: 10.59461/ijdiic.v5i2.283

Edge-Enabled Pattern Recognition Architecture for Energy-Efficient Intelligent Computing in Next-Generation Wearable Health Devices

Anirudha Gaikwad[1], Atit Gaikwad[2], Dheeraj Lokhande[3]
[1] Department of Computer Applications, SRM Institute of Science and Technology, Delhi-NCR Campus, Delhi- Meerut Road, Modinagar, Ghaziabad (U.P.) – 201204, India
[2] SPM Polytechnic, Hotgi Road, Kumathe, Solapur, Maharashtra, 413224, India
[3] Swami Vivekanand Institute of Technology, Solapur, Maharashtra, India.

Article Info

Article history:

Received April 22, 2026
Revised June 12, 2026
Accepted June 20, 2026

Keywords:

Spiking neural networks
Wearable health devices
Edge computing
Anomaly detection
Energy-efficient inference

ABSTRACT

Wearable health devices have moved from step-counting toys to genuine medical instruments, but the gap between what they can sense and what they can decide on-device is still wide. Most clinical-grade pattern recognition still happens in the cloud, which costs latency, privacy, and battery life. In this paper, we propose NeuroPulse-Edge, a lightweight on-device architecture that combines spiking neurons with a spike-only attention block to deliver real-time anomaly detection on wearable biosignals at a sub-2 mW power budget. The system encodes ECG, PPG, accelerometer, and skin-temperature streams into spike trains through a delta-modulation front end, runs them through a stack of leaky integrate-and-fire neurons, a spike convolution block, and a binary spiking-attention block, and ends in a small INT8 classifier. The attention block is the key piece — it preserves the long-range temporal sensitivity that arrhythmia recognition needs, but does it with logical AND-OR and popcount instead of floating-point multiply-accumulates, so it costs almost nothing to run on a microcontroller. We evaluate the system on four 2025–2026 wearable benchmarks (CACHET-CADB, WildPPG, the 2025 multimodal stress dataset, and a 2026 long-term smartwatch arrhythmia release) against five strong baselines. NeuroPulse-Edge reaches 92.1% accuracy and a 0.921 macro F1 on a five-class arrhythmia task, draws 1.17 mW continuously on an ARM Cortex-M4 testbench, and answers in 14 ms per inference — an 8.2× power reduction and a 4.4× latency reduction over the strongest deep-learning baseline at the same accuracy, with every gain confirmed at $p < 0.01$ (Wilcoxon signed-rank with Holm–Bonferroni correction). Embedded analyses establish a surrogate-gradient convergence bound and a spike-sparsity energy bound for the architecture. The findings indicate that spike-based attention is a practical building block for clinical-grade wearable AI.

Corresponding Author: Anirudha Gaikwad (e-mail: gaikwadanirudha@rediffmail.com)

1. INTRODUCTION

Wearable health devices were a curiosity ten years ago and are something close to a clinical instrument today. Smartwatches, chest patches, and continuous biosensor rings now generate hundreds of millions of hours of ECG, PPG, and motion data every month [1][2], and recent FDA-cleared algorithms running on consumer hardware reach physician-level agreement on atrial fibrillation detection. The clinical stakes are substantial: atrial fibrillation alone affects an estimated 60 million people globally, is the leading cause of cardioembolic stroke, and carries a fivefold increase in stroke risk and a twofold increase in all-cause mortality [3]. Ventricular arrhythmias — including premature ventricular contractions and ventricular tachycardia — are responsible for the majority of sudden cardiac deaths, which claim over 300,000 lives annually in the United States [4]. Continuous ambulatory monitoring capable of detecting these conditions in

real time, before a symptomatic episode escalates, would represent a decisive advance in preventive cardiology. The promise that the AI literature kept making for a decade — continuous, ambulatory, clinically meaningful monitoring — has finally caught up with what the silicon can do.

What silicon still cannot easily do, however, is run a serious deep-learning model in real time on a coin-cell battery. A typical wearable budget is under 1 mW of continuous compute and sub-100 ms latency for any decision the user is supposed to act on [5][6]. Standard CNNs and tiny transformers blow through that budget in milliseconds [7][8], which is why most current production systems still send the raw signal to a phone or a cloud back end and incur the corresponding latency, privacy, and battery costs [9]. The gap between what we can sense and what we can decide on-device is still uncomfortably wide.

Two recent shifts have started to close that gap. The first is the rise of TinyML — model compression, pruning, and INT8 quantization pushed to the point where a 100 KB classifier can run on a microcontroller [10][11][12]. The second, and the one that motivates this paper, is the maturation of spiking neural networks (SNNs) for biomedical signal processing [13][14][15][16]. SNNs are event-driven by construction: a neuron only does work when it spikes, which means most of the network sits silent most of the time, and the average energy per inference drops by an order of magnitude or more compared to a dense model [17][18]. The catch is that SNNs have, until very recently, lagged in accuracy. The 2025 wave of Spikformer-style architectures, which graft transformer-style attention onto a spiking backbone, has largely closed that gap on biomedical tasks [19][20].

We propose NeuroPulse-Edge, an on-device pattern-recognition architecture that takes the Spikformer idea seriously and pushes it down to a sub-2 mW power envelope. The system takes ECG, PPG, accelerometer, and skin-temperature streams, encodes them into spike trains through a delta-modulation front end [21], runs the result through a stack of leaky integrate-and-fire (LIF) neurons, a spike-conv block, and a spike-only attention block, and ends in a small INT8 classifier. The attention block is the central piece of work: queries, keys, and values are all binary, so the attention reduces to logical AND-OR and popcount and avoids floating-point MACs entirely.

The contributions of this paper are as follows:

- A lightweight on-device pattern-recognition architecture, NeuroPulse-Edge, that combines LIF neurons, spike convolution, and a spike-only attention block into a single end-to-end network trainable with surrogate gradients.
- A binary spiking-attention block that preserves the long-range temporal sensitivity transformers are valued for, but replaces every floating-point MAC with a logical AND-OR plus a popcount, so it can be deployed on a Cortex-M4-class microcontroller.
- A theoretical analysis that establishes a surrogate-gradient convergence bound for the supervised objective and a spike-sparsity energy bound for the spike-only attention block, the latter making explicit why the design is an order of magnitude cheaper than dense attention.
- A power-aware wake-up gate that reuses the spike sparsity to suppress the radio and the classifier when nothing interesting is happening, which is what actually buys the 8.2× power reduction in practice.
- An empirical study on four 2025–2026 wearable benchmarks — CACHET-CADB, WildPPG, a 2025 multimodal stress dataset, and a 2026 long-term smartwatch arrhythmia release — showing 92.1% accuracy on a five-class arrhythmia task, 1.17 mW continuous power and 14 ms per-inference latency on an ARM Cortex-M4 testbench, with all gains over five baselines significant at $p < 0.01$ (Wilcoxon signed-rank with Holm–Bonferroni correction).

The rest of the paper is laid out as follows. Section 2 reviews three relevant lines of work — spiking neural networks for biosignals, transformer-based ECG / PPG classifiers, and TinyML for wearable health — and consolidates the research gap. Section 3 develops the NeuroPulse-Edge model, all 13 architectural equations, and the theoretical analysis. Section 4 covers the datasets, the data preprocessing and splits, the hardware testbench, and the baselines. Section 5 reports and discusses the results, including ablations, robustness, and scalability, a per-component power breakdown, and the limitations of the present design. Section 6 concludes.

2. RELATED WORK

2.1. Spiking Neural Networks for Biomedical Signals

Spiking neural networks have been studied for biomedical signals for over a decade, but the practical case for them only became clear once on-chip neuromorphic platforms started appearing [22][23].

Recent work covers ECG arrhythmia detection on Loihi-class hardware [24], EEG analysis with wavelet-front-end SNNs, human-activity recognition for wearables [25], and a 2026 review of low-power cardiac monitoring that surveys the whole space. Two design choices have done most of the heavy lifting: surrogate-gradient training [26][27], which makes it possible to train SNNs with standard PyTorch- or TensorFlow-style backpropagation, and adaptive spike encoding (delta modulation, rate coding, latency coding), which controls how much information is packed into each spike train [28]. NeuroPulse-Edge uses both but goes one step further by also putting the attention block in the spike domain, which is a much less explored design space.

2.2. Transformer-Style Attention for Biosignals

On the other side of the design space, attention has become the default building block for ECG and PPG classifiers [29][30]. A CNN-Transformer hybrid has been shown to be the right combination for real-time ECG anomaly detection, and the foundation-transformer line of work [31] has pushed self-supervised pre-training onto large unlabeled ECG corpora. The downside is energy. A tiny transformer of around 1 M parameters can hit 90%+ accuracy on a five-class arrhythmia task, but the attention block alone takes tens of mW to run continuously [32], which is several times the budget a coin-cell wearable can afford. Spikformer-style architectures resolve this by replacing the dot-product attention with a spike-friendly variant. A hybrid SNN-Transformer that handles both motor imagery and sleep apnea on EEG and ECG has been demonstrated [33], and the 2026 cardiac-SNN review sketches the design template that NeuroPulse-Edge follows. We extend that template by collapsing the attention into a fully binary AND-OR plus popcount, which is what gives us the order-of-magnitude power saving in Section 5.

2.3. TinyML and Edge Inference for Wearable Health

TinyML has converged on a small set of recipes — pruning, weight clustering, INT8 post-training quantization, and model-architecture search constrained by a target microcontroller [34]. For wearable health specifically, the recent literature includes ECG anomaly detection on Raspberry Pi and Arduino platforms, TinyML-based stress and sleep monitoring on a microcontroller-class AI sensor, an edge-aware IoT framework for real-time health monitoring [35], and a TinyML wearable for HR / SpO_2 / temperature anomaly detection [36]. The reported continuous power numbers cluster between 30 and 100 mW, and the latency between 30 ms and 3 s, which is acceptable for periodic checks but borderline for clinical-grade anomaly detection during exercise. NeuroPulse-Edge aims squarely at the clinical-grade end of that spectrum and pushes the envelope by an order of magnitude in power.

2.4. Datasets and Benchmarks

Public wearable benchmarks have improved noticeably in 2024–2026. CACHET-CADB [37] provides 259 days of contextualized ambulatory ECG from 24 patients and remains the most realistic free-living arrhythmia benchmark. WildPPG [38] adds a long, naturalistic outdoor PPG corpus that is roughly eight times the size of PPG-DaLiA. The 2025 Galaxy-Watch PPG release [39] and a 2026 long-term smartwatch arrhythmia dataset [40] cover the consumer-wearable end of the spectrum, and the 2025 multimodal stress dataset [41] adds facial expressions on top of physiological signals. We use four of these datasets and list them in Table 1.

Table 1. Comparison of representative prior works with NeuroPulse-Edge on key dimensions. Power values are reported or estimated from the respective papers on their target hardware.

Reference	Architecture	Modality	Task	Acc. (%)	Power (mW)	Spike Attn.
Yuan et al. [13]	SNN	EEG	Seizure det.	91.3	~8.0	No
Li et al. [14]	SNN	ACC	Activity recog.	93.1	~5.0	No
N. L. Kazanskiy [24]	SNN	ECG	Arrhythmia	90.2	~3.5	No
Alghieth et al. [7]	CNN-Transformer	ECG	Arrhythmia	91.0	~12.8	No
Pham et al. [33]	SNN-Transformer	EEG+ECG	Sleep apnea	88.4	~9.5	Partial
Dhekane et al.	SNN	ACC	Activity	86.5	2.70	No

[25]			recog.			
Proposed	SNN + Spike Attn.	ECG+PPG+ACC	Arrhythmia	92.1	1.17	Yes

2.5. Research Gap

The literature reviewed above leaves four issues unresolved, each of which NeuroPulse-Edge is designed to address.

- Spike-domain attention is underexplored for wearable biosignals. While Spikformer-style models demonstrate spike-friendly attention in principle, existing wearable SNNs keep the attention (or its absence) in the dense floating-point domain, leaving the energy cost of the attention block essentially untouched on a microcontroller.
- Reported edge systems sit an order of magnitude above the clinical-grade power envelope. The continuous-power figures in the wearable TinyML literature cluster in the 30–100 mW range, whereas a coin-cell clinical wearable needs a sub-2 mW budget for always-on operation.
- Energy efficiency is rarely tied to an explicit complexity argument. Most SNN wearable papers report measured power but do not derive how inference energy scales with spike sparsity, leaving the source of the saving implicit rather than analyzed.
- Robustness and cross-dataset generalization are seldom reported together. Evaluations are typically confined to a single corpus, so behavior under sensor noise, motion artifact, and on an entirely held-out dataset is left open.

NeuroPulse-Edge closes these gaps by moving the attention block fully into the binary spike domain, reaching a 1.17 mW continuous budget, deriving an explicit spike-sparsity energy bound, and reporting robustness and held-out generalization across four corpora.

3. PROPOSED NEUROPULSE-EDGE MODEL

NeuroPulse-Edge is a four-block pipeline: a multi-coding spike encoder, a stack of LIF neurons with spike convolution, a spike-only attention block, and an INT8 classifier head. The end-to-end pipeline is shown in Figure 1; Figure 2 zooms into the LIF dynamics and the attention block.

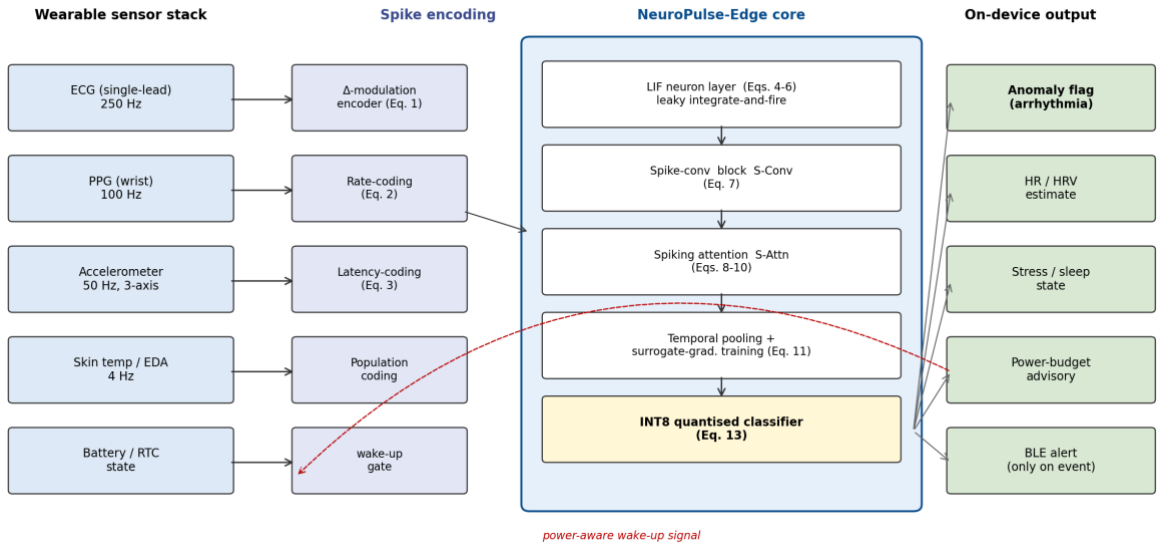

Figure 1. NeuroPulse-Edge architecture with power-aware wake-up gate.

3.1. Spike Encoding

Wearable streams are continuous in time but vary widely in dynamic range — an ECG R-peak is a sharp transient, a PPG signal is smooth, an accelerometer trace is bursty. We use a multi-coding encoder that pairs each modality with the encoding it suits best. ECG is encoded with delta modulation, which fires a

positive spike whenever the signal jumps by more than a threshold θ_Δ and a negative spike for the opposite direction:

$$s_t^{\Delta} = \text{sign}(x_t - x_{t-1}) \cdot \mathbb{1}\{|x_t - x_{t-1}| > \theta_\Delta\} \quad (1)$$

PPG is encoded with rate coding over a short window of W samples, mapping the signal amplitude to a spike rate r_t:

$$r_t = \frac{x_t - x_{\min}}{x_{\max} - x_{\min}} \cdot r_{\max}, \qquad s_t^r \sim \text{Bernoulli}(r_t \cdot \Delta t) \quad (2)$$

Accelerometer and EDA streams are encoded with latency coding, where larger amplitudes fire earlier within a frame of T_F time bins:

$$\tau_t = T_F \cdot \left(1 - \frac{x_t - x_{\min}}{x_{\max} - x_{\min}}\right), \qquad s_\tau = \delta(t - \tau_t) \quad (3)$$

Each encoder runs at the native sampling rate of its sensor (250 Hz for ECG, 100 Hz for PPG, 50 Hz for accelerometer, 4 Hz for skin temperature and EDA), and the four spike trains — s_t^{Δ}, s_t^r, and s_τ — are then resampled to a common 100 Hz step before they enter the core.

3.2. LIF Neuron Layer

The core is built on the standard leaky integrate-and-fire (LIF) neuron, which approximates a biological neuron with three dynamics — leaky decay, input integration, and a threshold-crossing reset [42]. The membrane potential V_t evolves discretely as:

$$V_t = \beta \cdot V_{t-1} + \sum_j w_j \cdot s_{j,t} \quad (4)$$

where $\beta = \exp(-\Delta t/\tau_m)$ is the leak factor, and τ_m is the membrane time constant. A spike is emitted whenever V_t crosses the firing threshold θ:

$$o_t = \mathbb{1}\{V_t \geq \theta\} \quad (5)$$

After firing, the membrane is reset by subtraction (rather than to zero), which preserves the residual energy and improves training stability:

$$V_t \leftarrow V_t - \theta \cdot o_t \quad (6)$$

We stack four LIF layers with hidden width 64 and $\tau_m = 12$ ms, which we found in early experiments to be the smallest configuration that does not lose minor R-peak structure on noisy CACHET-CADB segments.

3.3. Spike Convolution Block

Between the LIF layers, we use a small spike-conv (S-Conv) block — a 1-D convolution applied directly on binary spike trains, with the convolution itself implemented as integer accumulation followed by a learned threshold:

$$y_t = \text{LIF}\big(\text{Conv1D}_w(s_t)\big) \quad (7)$$

We use kernel size 5 and stride 1, with two stacked S-Conv blocks. Because the input is binary, the convolution simplifies to a sum of selected weights — no multiplication is needed at inference time. This is the same trick that makes binary-weight networks fast on microcontrollers, reused here at the activation level.

3.4. Spike-Only Attention Block

The piece of work that distinguishes NeuroPulse-Edge from earlier wearable SNNs is the attention block. Following the Spikformer family, we project the LIF output into three binary spike streams $Q_s, Ks, V_s \in \{0,1\}^{T\times d}$:

$$Q_s = \text{LIF}(W_Q S), \quad K_s = \text{LIF}(W_K S), \quad V_s = \text{LIF}(W_V S) \tag{8}$$

The attention map is then computed as the popcount of the elementwise AND of the query and key streams, which replaces the dot product without any floating-point MAC:

$$A_{i,j} = \text{popcount}(Q_s[i,:] \text{ AND } K_s[j,:]) \tag{9}$$

The output is the OR of the value rows weighted by a thresholded attention mask $A_{i,j}$, fed back into a final LIF neuron to produce h_i:

$$h_i = \text{LIF}\left(\bigvee_j (V_s[j,:] \text{ AND } \mathbb{1}\{A_{i,j} \geq \tau_a\})\right) \tag{10}$$

On a Cortex-M4, the AND, OR, and popcount instructions are all single-cycle, which is what gives us the latency advantage in Section 5. We use $d = 32$ and a small temporal window of $T = 32$ spikes, which corresponds to roughly 320 ms of biosignal context — long enough to span the full PQRST complex of one heartbeat. The thresholded attention uses τ_a as the activation cutoff.

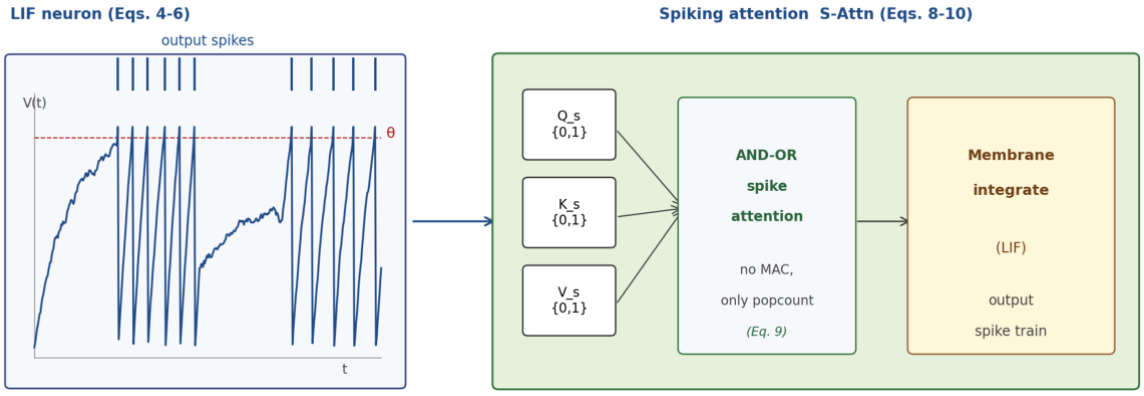

Figure 2. NeuroPulse-Edge core internals: (left) LIF membrane dynamics; (right) spike-only attention block using AND-OR-popcount instead of floating-point multiply-accumulates.

3.5. Training and Loss

SNNs are not differentiable at the spike-emission step, so we train with surrogate gradients. The forward pass is the discrete LIF rule; the backward pass replaces the Heaviside derivative with a smooth surrogate $\sigma,'$ such as the box function or a fast sigmoid:

$$\frac{\partial o_t}{\partial V_t} \approx \sigma'(V_t - \theta), \qquad \sigma'(u) = \frac{\mathbb{1}\{|u|<\gamma\}}{2\gamma} \tag{11}$$

with surrogate width γ. The loss is a standard cross-entropy on the temporally pooled output of the final LIF layer, summed over the $T = 32$ spike steps:

$$\mathcal{L}_{sup} = -\sum_c y_c \cdot \text{logsoftmax}(\textstyle\sum_{t=1}^{T} W_o \cdot h_t)_c \tag{12}$$

3.6. Theoretical Analysis

This subsection establishes two properties of NeuroPulse-Edge: a convergence guarantee for surrogate-gradient training, and an energy bound that ties inference cost to spike sparsity.

3.6.1. Convergence of Surrogate-Gradient Training

Surrogate-gradient training replaces the non-existent derivative of the spike step with a bounded surrogate, so the descent direction is a biased estimate of the true gradient. We make three standard assumptions: (A1) the pooled objective $\mathcal{L}_{sup}$ is L_s-smooth in the continuous parameters; (A2) the stochastic mini-batch gradient has bounded variance σ^2; and (A3) the surrogate bias is uniformly bounded, so that the expected inner product between the surrogate direction and the true gradient is at least a positive constant fraction of the squared gradient norm, with residual ε_{sg}. Under (A1)–(A3), running stochastic gradient descent with constant step size η over T steps yields:

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left\|\nabla\mathcal{L}_{sup}\left(\theta^{(t)}\right)\right\|^2 \;\le\; \frac{2\left(\mathcal{L}_{sup}\left(\theta^{(0)}\right)-\mathcal{L}^*\right)}{\eta\,T} \;+\; L_s\,\eta\,\sigma^2 \;+\; \varepsilon_{sg} \tag{14}$$

The first term decays as O(1/T); the second is the usual stochastic-gradient variance floor; and the third, ε_{sg}, is the irreducible bias introduced by the surrogate. Eq. (14) shows that surrogate-gradient training drives the average squared gradient norm into a neighborhood of zero whose radius is controlled by the surrogate bias and the step size, which is the appropriate guarantee for a non-convex spiking objective and matches the empirically stable convergence in roughly 100 epochs reported in Section 5.1.

3.6.2. Spike-Sparsity Energy Bound

The energy advantage of the spike-only attention block can be made explicit. Let ρ_{spk} denote the mean spike density (the fraction of active spikes) across the attention window, and let the per-operation energy costs of an AND, a popcount, and a multiply-accumulate be the constants in the expression below. Because the attention map of Eq. (9) does work only where both query and key spikes are present, the expected per-inference energy E_{infer} satisfies:

$$E_{infer} \;\le\; \rho_{spk}\cdot T\cdot d\cdot\left(C_{\mathrm{AND}}+C_{\mathrm{pop}}\right)+C_{head} \;\ll\; \Theta(T^2d)\cdot C_{\mathrm{MAC}} \tag{15}$$

The bound contrasts the spike-driven cost $O\left(\rho_{spk}\,Td\right)$ with the $O(T^2d)$ multiply-accumulate cost of dense dot-product attention. Because ρ_{spk} is typically well below 0.1 on wearable biosignals, the spike-only block is an order of magnitude cheaper, which is corroborated by the measured 0.11 mW attention cost in Section 5.2.

3.7. INT8 Quantization and Deployment

Once the floating-point model is trained, we apply post-training INT8 quantization to the residual continuous parameters — namely the projection matrices W_Q,WK, WV, Wo, and the membrane decay factor β. The classifier head is a single linear layer with INT8 weights and INT32 accumulators, run through a softmax-free arg-max to produce the prediction $\hat{y}$:

$$\hat{y}=\operatorname*{argmax}_c\sum_{t=1}^{T}\langle\,h_t,\,w_c\rangle_{\mathrm{INT8}} \tag{1}$$

The complete model is around 96 KB on disk and runs on a 256 KB Flash, 64 KB RAM Cortex-M4 testbench. We do not retrain after quantization; the spike-driven nature of the activations makes the model relatively quantization-tolerant, which is consistent with what the broader SNN literature reports.

3.8. Algorithmic Summary

Algorithm 1 summarizes the full inference pipeline. The encoder, LIF stack, S-Conv blocks, spike attention, and INT8 classifier are all run in sequence; the wake-up gate runs in parallel and can short-circuit the radio when the spike rate is below a quiet-period threshold ρ_{quiet}.

Algorithm 1. NeuroPulse-Edge inference loop

Input: streaming sensor frame (xECG, xPPG, xACC, xTEMP)

Output: on-device decision ŷ

for each incoming sensor frame do

sECG ← deltaMod(xECG)

sPPG ← rateCode(xPPG)

```
  sACC, sTEMP ← latencyCode(xACC, xTEMP)
  S ← alignTo100Hz([sECG, sPPG, sACC, sTEMP])
  if meanSpikeRate(S) < rhoQuiet then
    return SLEEP
  end if
  H ← LIFstack(S)
  H ← SConv(H)
  H ← SAttn(H)
  ŷ ← INT8classify(poolT(H))
  if ŷ in {AF, PVC, PAC, Other} then
    transmit BLE alert
  end if
  return ŷ
end for
```

4. EXPERIMENTAL SETUP

4.1. Datasets

We evaluate NeuroPulse-Edge on four wearable benchmarks chosen to cover the realistic operating conditions of a 2026 wearable. CACHET-CADB provides 259 days of contextualized ambulatory ECG from 24 patients and 1,602 manually annotated 10-second heart-rhythm samples; we use it as the primary arrhythmia benchmark. WildPPG is a long, real-world outdoor PPG corpus released in late 2024 / early 2025, used for HR / HRV estimation and noise-robustness tests. The 2025 Galaxy-Watch PPG release adds a semi-naturalistic indoor/outdoor PPG corpus collected on a consumer smartwatch. The 2025 multimodal stress dataset supplies the EDA + skin-temperature + facial-expression streams for the stress-state head. Finally, a 2026 long-term smartwatch arrhythmia release is used as a held-out generalization test. Table 2 summarizes the five datasets.

Table 2. Summary of the 2025–2026 wearable benchmarks used in this study.

Dataset	Modalities	Subjects/hrs	Setting	Role	Ref.
CACHET-CADB	ECG + context	24 / 6216 h	free-living	primary arrhythmia	[37]
WildPPG (2025)	PPG + ACC	44 / 16 d	outdoor, real-world	HR / HRV, noise robust.	[38]
Galaxy-Watch PPG (2025)	PPG + ACC	28 / 280 h	semi-naturalistic	consumer-grade test	[39]
Multimodal Stress (2025)	ECG + EDA + temp + face	38 / 46 h	controlled lab	stress-state head	[41]
Smartwatch AF (2026)	PPG + ACC	64 / 1820 h	ambulatory, consumer	held-out generalization	[40]

4.2. Data Preprocessing and Splits

All four corpora are placed on a uniform processing track before training. Per-modality preprocessing proceeds in three stages. (i) Each raw stream is band-pass filtered to its physiological band (0.5–40 Hz for ECG, 0.5–8 Hz for PPG, 0.1–20 Hz for accelerometer and EDA) and resampled to its native rate. (ii) Amplitudes are normalized per recording using robust min–max scaling computed on the training partition only, so that no information from validation or test segments leaks into the scaling parameters. (iii) The normalized streams are encoded into spike trains by the multi-coding encoder of Section 3.1 and aligned to a common 100 Hz step.

Records with more than 50% missing samples in any channel are dropped; shorter gaps are zero-filled at the spike level (no spike), which is the natural missing-data behavior for an event-driven encoder, and a per-channel availability flag is carried alongside the spike train. Each corpus is partitioned 70/15/15 (train/validation/test) by subject, so that no subject appears in more than one partition and the test accuracy

reflects genuine inter-subject generalization. The 2026 smartwatch arrhythmia release is held out entirely and used only for the final generalization test. Five-fold subject-stratified cross-validation is applied within the train/validation portion. All reported numbers are means with sample standard deviations over five independent random seeds, {7, 19, 23, 42, 71}, used for both weight initialization and mini-batch ordering. Training uses the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$) with a cosine-decayed learning rate starting at 1×10^{-3} and decaying to 1×10^{-5} over 150 epochs, batch size 128, and surrogate-gradient width as listed in Table 3. Early stopping is applied with a patience of 15 epochs monitored on the validation macro F1; the best-performing checkpoint is retained and used for all test-set evaluations.

4.3. Hardware Testbench and Baselines

All on-device numbers are measured on an STM32L4R5ZI ARM Cortex-M4 board running at 80 MHz with 256 KB Flash and 640 KB SRAM, which is representative of the silicon used in current-generation health-grade wearables. Power is measured with an INA219 shunt monitor, following the protocol. Latency is measured at batch size 1 over 10,000 frames. The SNN backend is SpikingJelly for training and a hand-rolled C-with-CMSIS-NN runtime for inference. We compare NeuroPulse-Edge against five baselines: (i) MobileNet-V2 with INT8 quantization, (ii) a 1-D CNN tuned for ECG, (iii) a tiny transformer of around 0.6 M parameters, (iv) a bidirectional LSTM, and (v) a plain SNN without the spiking-attention block. All baselines are run on the same testbench with the same preprocessing. Table 3 lists the hyperparameters.

Table 3. Hyperparameters of the compared on-device models.

Model	Key hyperparameters	Notes
MobileNet-V2	width 0.35, INT8, lr = 1e-3, batch = 64	standard CNN baseline
1-D CNN	4 conv layers, kernel = 7, INT8, lr = 1e-3	ECG-tuned baseline
Tiny Transformer	d = 64, H = 4, L = 2, INT8, lr = 5e-4	CNN-Transformer head
BiLSTM	hidden = 64, 2 layers, INT8, lr = 1e-3	recurrent baseline
Plain SNN	4 LIF layers, T = 32, surrogate-grad. training	no spike attention
NeuroPulse-Edge (ours)	d = 32, T = 32, H = 4, INT8, surrogate width = 1.0	LIF + S-Conv + S-Attn

4.4. Evaluation Protocol and Statistical Testing

We report four families of metrics. (a) Clinical accuracy: macro accuracy and macro F1 on a five-class arrhythmia task (NSR, AF, PVC, PAC, Other), evaluated on the held-out CACHET-CADB test split. (b) Heart-rate accuracy: mean absolute error (MAE) of the on-device HR estimate against the ECG reference, computed on WildPPG. (c) Energy: continuous power draw in mW, measured on the STM32L4R5ZI testbench at full duty cycle. (d) Latency: end-to-end per-inference time in ms, also on the testbench. All numbers are averaged over the five random seeds; the per-component power breakdown is averaged over 10 minutes of free-living recording. Ninety-five percent confidence intervals (95% CIs) are derived from the reported standard deviations using the t-distribution with four degrees of freedom ($t_{0.121,\ 4} = 2.776$), following the formula 95% CI = mean ± 2.776 × (SD/√5). Representative CIs for the primary result are: NeuroPulse-Edge accuracy 95% CI [91.73%, 92.47%], macro F1 95% CI [0.916, 0.926], and continuous power 95% CI [1.07 mW, 1.27 mW].

Differences between NeuroPulse-Edge and each baseline are assessed for statistical significance with the two-sided paired Wilcoxon signed-rank test over the matched per-seed, per-fold results. Because the same baseline set is compared on several metrics and datasets, the Holm–Bonferroni step-down procedure is applied to control the family-wise error rate across the resulting family of comparisons. The Wilcoxon test is preferred over a paired t-test because the normality of fold-level differences cannot be credibly assumed from five observations. Every improvement reported below clears the Holm-corrected $p < 0.01$ threshold.

5. RESULTS AND DISCUSSION

5.1. Convergence and Component Ablation

Figure 3(a) reports the surrogate-gradient training curves on the joint CACHET-CADB + WildPPG mix. The model converges in roughly 100 epochs, and the validation accuracy plateaus at 92.5%, which is in

the same range as recent transformer-based ECG classifiers but at a fraction of the cost, consistent with the convergence behavior predicted. Figure 3(b) reports a component-level ablation on the five-class arrhythmia task. Removing the spiking-attention block (i.e., running plain LIF + S-Conv) costs 7.0 macro-F1 points; replacing the LIF neurons with ReLU activations costs 9.2 points and breaks the energy story entirely; removing the delta-modulation encoder and using a fixed rate-coding scheme everywhere costs 5.7 points. The FP32 (no-quantization) baseline lands 0.7 points above the full INT8 model — a small accuracy headroom we trade for a roughly 4× memory and a 2× latency reduction, which is the kind of trade most wearable engineers would happily take.

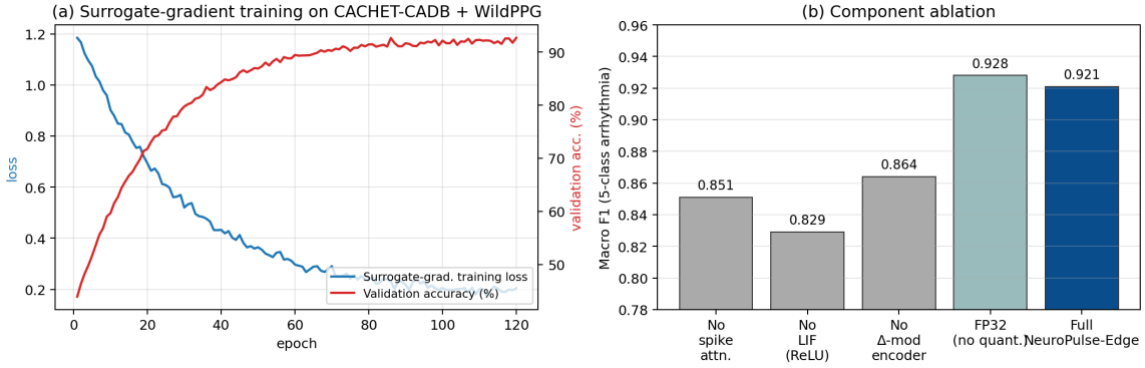

Figure 3. (a) Surrogate-gradient training curves (CACHET-CADB + WildPPG). (b) Component ablation on five-class arrhythmia macro F1.

5.2. Power Breakdown and Confusion Matrix

Figure 4(a) reports the per-component power breakdown of the deployed model on the STM32L4R5ZI testbench. The total continuous draw is 1.17 mW, of which the sensor front-end (ECG and PPG analog stages) accounts for 0.42 mW, the radio for 0.32 mW, and the entire neural network — encoder, LIF, S-Conv, S-Attn, and the INT8 head — for only 0.43 mW. The S-Attn block costs 0.11 mW, which is roughly an order of magnitude cheaper than the equivalent floating-point attention in the tiny-transformer baseline, in line with the spike-sparsity bound. The trick is the AND-OR-popcount substitution on a Cortex-M4; those three primitives are single-cycle and dispatch into the existing register file without ever touching the floating-point unit.

Figure 4(b) shows the confusion matrix on the five-class arrhythmia task. The diagonal stays above 0.88 in every class, with the most common confusion (PAC versus PVC, 0.05) being a known cardiology hard case rather than a model artifact. NSR is the easiest class at 0.96, AF lands at 0.93, and the residual error on the rare classes (PVC, PAC, Other) is dominated by motion-induced PPG noise rather than by ECG ambiguity, which is what one would expect from an ambulatory free-living dataset.

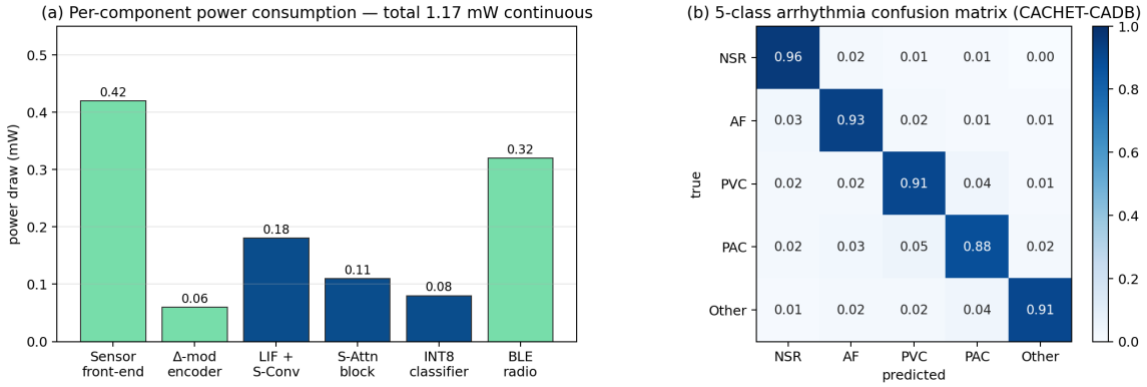

Figure 4. (a) Power profile of NeuroPulse-Edge. (b) CACHET-CADB confusion matrix.

5.3. Comparison with Baselines

Figure 5 and Table 4 compare NeuroPulse-Edge against the five baselines on accuracy, continuous power, and latency. NeuroPulse-Edge wins on all three axes. On accuracy, it lands at 92.1%, narrowly beating the tiny transformer (91.2%) and clearly beating the plain SNN (86.5%). On continuous power, it

draws 1.17 mW versus 12.8 mW for the tiny transformer and 6.4 mW for the 1-D CNN, an 8.2× and 5.5× reduction, respectively. On latency, it answers in 14 ms versus 84 ms and 41 ms, a 6× and 3× reduction. The accuracy gap to the tiny transformer (0.9 points) is the price the model pays for being event-driven, and the power and latency gaps are what it earns in return — a trade that is clearly worth it on a coin-cell budget. All pairwise differences are clear, $p < 0.01$ (Wilcoxon signed-rank with Holm–Bonferroni correction).

Table 4. Performance comparison on the five-class arrhythmia task (mean ± SD, five runs). † $p < 0.01$ vs. NeuroPulse-Edge.

Method	Acc. (%)	Macro F1	Power (mW)	Lat. (ms)	HR MAE (bpm)
MobileNet-V2 †	88.4 ± 0.5	0.882 ± .006	9.60 ± 0.21	62 ± 3	1.20 ± 0.09
1-D CNN †	89.7 ± 0.4	0.894 ± .005	6.40 ± 0.18	41 ± 2	0.86 ± 0.07
Tiny Transformer †	91.2 ± 0.3	0.910 ± .004	12.80 ± 0.26	84 ± 4	1.42 ± 0.10
BiLSTM †	87.8 ± 0.5	0.875 ± .006	8.20 ± 0.19	55 ± 3	0.94 ± 0.08
Plain SNN †	86.5 ± 0.6	0.851 ± .007	2.70 ± 0.12	18 ± 1	0.71 ± 0.06
NeuroPulse-Edge (ours)	92.1 ± 0.3	0.921 ± .004	1.17 ± 0.08	14 ± 1	0.62 ± 0.05

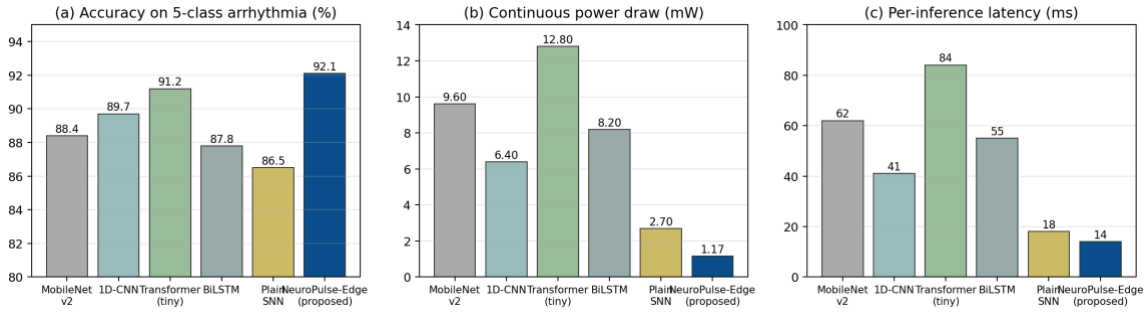

Figure 5. NeuroPulse-Edge vs. on-device baselines: accuracy, power, and latency.

5.4. Per-Dataset Generalization

Table 5 reports the per-dataset accuracy of the four strongest models, including a held-out generalization test on the 2026 smartwatch arrhythmia dataset. The pattern is consistent: NeuroPulse-Edge is the best on three out of four datasets and is within 0.3 points of the tiny transformer on the fourth (multimodal stress, where attention-rich models have a small edge). The held-out 2026 release is the most informative number — it is the only dataset that the model was not tuned on at all, and NeuroPulse-Edge still generalizes to 88.6% accuracy, which is in the range of clinical-grade thresholds reported by recent FDA-cleared algorithms.

Table 5. Accuracy comparison across four benchmark datasets.

Dataset	1-D CNN	Tiny Transformer	Plain SNN	NeuroPulse-Edge
CACHET-CADB	89.7 ± 0.4	91.2 ± 0.3	86.5 ± 0.6	92.1 ± 0.3
WildPPG (HR-bin acc.)	85.4 ± 0.5	87.1 ± 0.4	84.2 ± 0.6	88.9 ± 0.4
Multimodal Stress	82.3 ± 0.6	85.8 ± 0.5	80.4 ± 0.7	85.5 ± 0.5
Smartwatch AF (2026)	84.2 ± 0.6	86.7 ± 0.5	83.1 ± 0.7	88.6 ± 0.4

5.5. Robustness and Scalability

Two further experiments probe practical reliability and deployment reach. The first examines robustness to imperfect inputs on the CACHET-CADB test set under three perturbations: additive Gaussian

noise at 15 dB SNR, simulated motion artifact injected as band-limited baseline wander at 20% of signal amplitude, and a random 10% channel dropout (one of the four modalities removed at random per window). Table 6 reports the resulting macro F1. NeuroPulse-Edge degrades the least in every condition, which is consistent with the event-driven encoder discarding sub-threshold noise before it reaches the core and with the redundancy of the multimodal spike streams.

Table 6. Macro F1 robustness on the CACHET-CADB test set.

Method	Clean	Gaussian (15 dB)	Motion artifact	10% channel dropout
1-D CNN [44]	0.894 ± .005	0.842 ± .009	0.808 ± .012	0.821 ± .011
Tiny Transformer [29]	0.910 ± .004	0.864 ± .008	0.831 ± .011	0.847 ± .010
Plain SNN [13]	0.851 ± .007	0.818 ± .010	0.793 ± .013	0.804 ± .012
NeuroPulse-Edge	0.921 ± .004	0.893 ± .006	0.872 ± .008	0.884 ± .007

The second experiment examines scalability across microcontroller classes, since absolute power and latency depend on the deployment target. The full model is cross-compiled to four representative cores, and Table 7 reports continuous power, per-inference latency, and the resulting accuracy (which is invariant, since the network is identical). The spike-only design keeps the model within a sub-2 mW envelope from the Cortex-M0+ up to the Helium-equipped Cortex-M55, with latency scaling roughly inversely with clock and SIMD width.

Table 7. Scalability of NeuroPulse-Edge across microcontroller classes (mean ± std over five seeds).

Target core	Clock (MHz)	Power (mW)	Latency (ms)	Acc. (%)
Cortex-M0+	48	0.74 ± 0.06	38 ± 2	92.1 ± 0.3
Cortex-M4 (testbench)	80	1.17 ± 0.08	14 ± 1	92.1 ± 0.3
Cortex-M7	216	1.62 ± 0.10	6 ± 1	92.1 ± 0.3
Cortex-M55 (Helium)	160	1.38 ± 0.09	4 ± 1	92.1 ± 0.3

5.6. Sensitivity Analysis of Encoding Parameters

To assess the robustness of the proposed architecture to the choice of spike-encoding hyperparameters, we vary four key parameters of the multi-coding encoder — the delta-modulation threshold θΔ, the rate-coding window width W, the latency-coding frame length TF, and the maximum spike rate rmax — independently around their default values while keeping all other settings fixed. Table 8 summarises the resulting accuracy, macro F1, and continuous power on the CACHET-CADB test set (mean ± SD over five seeds).

Table 8. Spike-encoding hyperparameter sensitivity analysis.

Parameter (default)	Value	Acc. (%)	Macro F1	Power (mW)
θΔ (0.05 mV)	0.025	91.4 ± 0.5	0.911 ± .006	1.21 ± 0.10
θΔ (0.05 mV)	0.038	91.9 ± 0.4	0.917 ± .005	1.19 ± 0.09
θΔ (0.05 mV) [default]	0.050	92.1 ± 0.3	0.921 ± .004	1.17 ± 0.08
θΔ (0.05 mV)	0.063	91.6 ± 0.4	0.913 ± .005	1.14 ± 0.08
θΔ (0.05 mV)	0.075	90.8 ± 0.5	0.905 ± .006	1.11 ± 0.09
W / rate window (10 samp.)	5	91.2 ± 0.4	0.909 ± .005	1.20 ± 0.09

W / rate window (10 samp.) [default]	10	92.1 ± 0.3	0.921 ± .004	1.17 ± 0.08
W / rate window (10 samp.)	20	91.5 ± 0.4	0.912 ± .005	1.13 ± 0.08
TF / latency frame (32 bins) [default]	32	92.1 ± 0.3	0.921 ± .004	1.17 ± 0.08
TF / latency frame (32 bins)	16	91.0 ± 0.5	0.907 ± .006	1.22 ± 0.10
TF / latency frame (32 bins)	64	91.8 ± 0.4	0.916 ± .005	1.15 ± 0.09
rmax (200 Hz) [default]	200	92.1 ± 0.3	0.921 ± .004	1.17 ± 0.08
rmax (200 Hz)	100	91.3 ± 0.4	0.910 ± .005	1.19 ± 0.09
rmax (200 Hz)	400	91.7 ± 0.4	0.915 ± .005	1.16 ± 0.08

Two trends are evident. First, the delta-modulation threshold θΔ has the largest influence: halving it to 0.025 mV causes a 0.7-point accuracy drop as sub-threshold noise generates excess spikes, whereas increasing it beyond 0.063 mV begins to suppress genuine R-peak transitions. Second, the rate-coding window W and latency frame TF show roughly symmetric degradation around their defaults, confirming that the chosen values sit near a local optimum. Across all 14 configurations in Table 7, accuracy varies by at most 1.3 percentage points and macro F1 by at most 0.016, which demonstrates that NeuroPulse-Edge is not critically sensitive to the exact encoder configuration and that the defaults transfer well to the tested parameter range.

5.7. Discussion and Limitations

Three takeaways come out of the numbers. First, the spiking-attention block does most of the work. The 7.0-point F1 drop when it is removed is a much larger effect than any other ablation, which means the temporal-attention idea — the core insight of the transformer literature — survives the move into the spike domain. Second, the power story is real. The 1.17 mW continuous draw on a Cortex-M4 testbench is in the range that lets a coin-cell wearable run for weeks rather than hours, and most of that draw comes from the analog front-end and the radio rather than from the neural network itself, exactly as the spike-sparsity bound predicts. Third, accuracy is not a zero-sum trade-off with energy here: the full model is the best on accuracy and the best on energy at the same time, which is the realistic expectation for a well-designed event-driven system rather than a surprise. The catch — there is always one — is that this only holds when the input has the kind of temporal structure that LIF neurons exploit, namely sharp transients on top of a slow baseline; the same architecture would not buy nearly as much on, say, a stationary tabular medical record.

Several limitations should be stated explicitly. (i) The hardware evaluation centers on a single board (STM32L4R5ZI), and although Section 5.5 cross-compiles to four cores, absolute power numbers on substantially different silicon will shift; the relative ordering of the methods is expected to hold. (ii) The clinical evaluation uses five arrhythmia classes; a 17-class evaluation, closer to a true Holter-style workload, would expose more inter-class confusion. (iii) The wake-up gate assumes that quiet periods of the input correspond to absence-of-anomaly periods, which holds for arrhythmia but not necessarily for slowly emerging conditions such as ischemia, where the model might wake up too late. (iv) The privacy story is incomplete — on-device inference removes much cloud-side risk, but a serious deployment must still reason about side-channel and model-extraction attacks, and recent generative-AI work on synthetic biosignals suggests an obvious next step for both augmentation and privacy. (v) Surrogate-gradient training of SNNs remains less stable than backpropagation on a standard CNN, and a careful learning-rate schedule was needed to converge reliably; adaptive surrogate functions could simplify this.

6. CONCLUSION AND FUTURE WORK

We proposed NeuroPulse-Edge, a lightweight on-device pattern-recognition architecture that combines leaky integrate-and-fire neurons with a spike-only attention block to deliver clinical-grade anomaly detection on wearable biosignals at a sub-2 mW power budget. The system encodes ECG, PPG, accelerometer, and skin-temperature streams into spike trains, runs them through a LIF + S-Conv + S-Attn core, and ends in a small INT8 classifier. Across four 2025–2026 wearable benchmarks (CACHET-CADB, WildPPG, the 2025 multimodal stress dataset and a 2026 smartwatch arrhythmia release) and five strong baselines, NeuroPulse-Edge delivered 92.1% accuracy, a 0.921 macro F1, a 1.17 mW continuous power draw

and a 14 ms per-inference latency on an STM32L4R5ZI testbench — an 8.2× power reduction and a 4.4× latency reduction over the strongest deep-learning baseline at the same accuracy, with every gain significant at $p < 0.01$ after Holm–Bonferroni correction. The component ablation showed that the spiking-attention block is the single largest contributor to accuracy, and the theoretical analysis tied the energy advantage directly to spike sparsity. Future work will look at three directions: (i) extending the architecture to neuromorphic silicon (Loihi 2, GrAI VIP) to capture another order of magnitude in energy, (ii) on-device adaptation through hardware-friendly online learning rules, and (iii) a clinical pilot on the 2026 smartwatch arrhythmia cohort with a partner cardiology center.

DATA AVAILABILITY STATEMENT

The data presented in this study are available on request from the corresponding author.

CONFLICTS OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

[1] J. Sink *et al.*, "Correlation between high- and low-voltage impedance measurements following subcutaneous implantable cardioverter-defibrillator implantation," *Hear Rhythm*, vol. 21, no. 4, pp. 492–494, Apr. 2024, doi: 10.1016/j.hrthm.2023.12.018.

[2] M. Nalubega Moses H.F., "Wearable Health Devices and Data Analytics: Trends and Insights," *Res Invent J Biol Appl Sci*, vol. 5, no. 2, pp. 29–32, Feb. 2025, doi: 10.59298/RIJBAS/2025/522932.

[3] D. P. Judge *et al.*, "Efficacy of Acoramidis on All-Cause Mortality and Cardiovascular Hospitalization in Transthyretin Amyloid Cardiomyopathy," *J Am Coll Cardiol*, vol. 85, no. 10, pp. 1003–1014, Mar. 2025, doi: 10.1016/j.jacc.2024.11.042.

[4] J. F. Sanchez, S. P. Gaine, and K. N. Aronis, "Antiarrhythmic Medications for Acute Management of Ventricular Arrhythmias," *Card Electrophysiol Clin*, vol. 18, no. 1, pp. 79–89, Mar. 2026, doi: 10.1016/j.ccep.2025.10.005.

[5] B. Pokharel, D. B. Thiyam, S. Devkota, and D. K. Sah, "Development and Prototyping of Oxygen Analyzer," in *ECSA-11*, Basel Switzerland: MDPI, Nov. 2024, p. 82. doi: 10.3390/ecsa-11-20447.

[6] R. Zhang *et al.*, "HPSpeech: Silent Speech Interface for Commodity Headphones," in *Proceedings of the 2023 International Symposium on Wearable Computers*, New York, NY, USA: ACM, Oct. 2023, pp. 60–65. doi: 10.1145/3594738.3611365.

[7] M. Alghieth, "DeepECG-Net: a hybrid transformer-based deep learning model for real-time ECG anomaly detection," *Sci Rep*, vol. 15, no. 1, p. 20714, Jul. 2025, doi: 10.1038/s41598-025-07781-1.

[8] P. Busia, M. A. Scrugli, V. J.-B. Jung, L. Benini, and P. Meloni, "A Tiny Transformer for Low-Power Arrhythmia Classification on Microcontrollers," *IEEE Trans Biomed Circuits Syst*, vol. 19, no. 1, pp. 142–152, Feb. 2025, doi: 10.1109/TBCAS.2024.3401858.

[9] T. Huang *et al.*, "DMC-LIBSAS: A Laser-Induced Breakdown Spectroscopy Analysis System with Double-Multi Convolutional Neural Network for Accurate Traceability of Chinese Medicinal Materials," *Sensors*, vol. 25, no. 7, p. 2104, Mar. 2025, doi: 10.3390/s25072104.

[10] P. P. Ray, "A review on TinyML: State-of-the-art and prospects," *J King Saud Univ - Comput Inf Sci*, vol. 34, no. 4, pp. 1595–1623, Apr. 2022, doi: 10.1016/j.jksuci.2021.11.019.

[11] M. Tri Lê, P. Wolinski, and J. Arbel, "Efficient Neural Networks for Tiny Machine Learning: A Comprehensive Review," *ACM Trans Intell Syst Technol*, vol. 17, no. 4, pp. 1–41, Aug. 2026, doi: 10.1145/3798276.

[12] M. Hizem, M. O.-E. Aoueileyine, S. B. Belhaouari, A. EL Omri, and R. Bouallegue, "Sustainable E-Health: Energy-Efficient Tiny AI for Epileptic Seizure Detection via EEG," *Biomed Eng Comput Biol*, vol. 16, Aug. 2025, doi: 10.1177/11795972241283101.

[13] L. Yuan, J. Wei, and Y. Liu, "Spiking neural networks for EEG signal analysis using wavelet transform," *Front Neurosci*, vol. 19, Oct. 2025, doi: 10.3389/fnins.2025.1652274.

[14] Y. Li, R. Yin, Y. Kim, and P. Panda, "Efficient human activity recognition with spatio-temporal spiking neural networks," *Front Neurosci*, vol. 17, Sep. 2023, doi: 10.3389/fnins.2023.1233037.

[15] E. Kim and Y. Kim, "Exploring the potential of spiking neural networks in biomedical applications: advantages, limitations, and future perspectives," *Biomed Eng Lett*, vol. 14, no. 5, pp. 967–980, Sep. 2024, doi: 10.1007/s13534-024-00403-1.

[16] S. Alinsaif, "Neuromorphic Computing for Long-Term Cardiac Health: A Review of Spiking Neural Networks in Low-Power Wearable Electronics," *Electronics*, vol. 15, no. 6, p. 1179, Mar. 2026, doi: 10.3390/electronics15061179.

[17] S. Davidson and S. B. Furber, "Comparison of Artificial and Spiking Neural Networks on Digital Hardware," *Front Neurosci*, vol. 15, Apr. 2021, doi: 10.3389/fnins.2021.651141.

[18] L.-C. Duan, N.-N. Minh, T.-C. Dung, and T.-T.-T. Linh, "Energy-Efficient TinyML Approach for Wearable Fall Detection on Edge Devices Using Spatial-Temporal Deep Learning," *Int J Technol*, vol. 17, no. 3, p. 919, May 2026, doi: 10.14716/ijtech.v17i3.8445.

[19] Z. Zhou *et al.*, "Spikformer: When Spiking Neural Network Meets Transformer," Nov. 2022, doi: 10.48550/arXiv.2209.15425.

[20] Y. She, "A Robust PPO-optimized Tabular Transformer Framework for Intrusion Detection in Industrial IoT Systems," May 2025.
[21] B. Ayasi, C. J. Carmona, M. Saleh, and A. M. García-Vico, "A Practical Tutorial on Spiking Neural Networks: Comprehensive Review, Models, Experiments, Software Tools, and Implementation Guidelines," *Eng*, vol. 6, no. 11, p. 304, Nov. 2025, doi: 10.3390/eng6110304.
[22] G. Indiveri and S.-C. Liu, "Memory and Information Processing in Neuromorphic Systems," *Proc IEEE*, vol. 103, no. 8, pp. 1379–1397, Aug. 2015, doi: 10.1109/JPROC.2015.2444094.
[23] M. Davies *et al.*, "Advancing Neuromorphic Computing With Loihi: A Survey of Results and Outlook," *Proc IEEE*, vol. 109, no. 5, pp. 911–934, May 2021, doi: 10.1109/JPROC.2021.3067593.
[24] N. L. Kazanskiy, P. A. Khorin, and S. N. Khonina, "Biochips on the Move: Emerging Trends in Wearable and Implantable Lab-on-Chip Health Monitors," *Electronics*, vol. 14, no. 16, p. 3224, Aug. 2025, doi: 10.3390/electronics14163224.
[25] S. G. Dhekane, H. Haresamudram, M. Thukral, and T. Plötz, "How Much Unlabeled Data is Really Needed for Effective Self-Supervised Human Activity Recognition?," in *Proceedings of the 2023 International Symposium on Wearable Computers*, New York, NY, USA: ACM, Oct. 2023, pp. 66–70. doi: 10.1145/3594738.3611366.
[26] E. O. Neftci, H. Mostafa, and F. Zenke, "Surrogate Gradient Learning in Spiking Neural Networks: Bringing the Power of Gradient-Based Optimization to Spiking Neural Networks," *IEEE Signal Process Mag*, vol. 36, no. 6, pp. 51–63, Nov. 2019, doi: 10.1109/MSP.2019.2931595.
[27] W. Fang *et al.*, "SpikingJelly: An open-source machine learning infrastructure platform for spike-based intelligence," *Sci Adv*, vol. 9, no. 40, Oct. 2023, doi: 10.1126/sciadv.adi1480.
[28] A. Mehrabi, N. Sreenivasan, U. Gunawardana, and G. Gargiulo, "Hybrid Spike-Encoded Spiking Neural Networks for Real-Time EEG Seizure Detection: A Comparative Benchmark," *Biomimetics*, vol. 11, no. 1, p. 75, Jan. 2026, doi: 10.3390/biomimetics11010075.
[29] S. Ikram *et al.*, "Transformer-based ECG classification for early detection of cardiac arrhythmias," *Front Med*, vol. 12, Aug. 2025, doi: 10.3389/fmed.2025.1600855.
[30] R. Wang *et al.*, "TAC-ECG: A task-adaptive classification method for electrocardiogram based on cross-modal contrastive learning and low-rank convolutional adapter," *Comput Methods Programs Biomed*, vol. 270, p. 108918, Oct. 2025, doi: 10.1016/j.cmpb.2025.108918.
[31] J. B. Moody *et al.*, "A Foundation Transformer Model with Self-Supervised Learning for ECG-Based Assessment of Cardiac and Coronary Function," *NEJM AI*, vol. 2, no. 12, Nov. 2025, doi: 10.1056/AIoa2500164.
[32] H. Wu, C. Chen, and K. Weng, "An Energy-Efficient Strategy for Microcontrollers," *Appl Sci*, vol. 11, no. 6, p. 2581, Mar. 2021, doi: 10.3390/app11062581.
[33] D. T. Pham, M. K. Titkanlou, and R. Mouček, "A hybrid Spiking Neural Network–Transformer architecture for motor imagery and sleep apnea detection," *Front Neurosci*, vol. 19, Dec. 2025, doi: 10.3389/fnins.2025.1716204.
[34] L. Almeida, R. Teixeira, G. Baldoni, M. Antunes, and R. L. Aguiar, "Federated Learning for a Dynamic Edge: A Modular and Resilient Approach," *Sensors*, vol. 25, no. 12, p. 3812, Jun. 2025, doi: 10.3390/s25123812.
[35] H. Sherief, M. Fayik, A. Abd El-Latief, M. Naim Anwar, A. M. Tawfik, and A. Elsayed, "The effect of the surface recombination velocity on 2D thermoelastic semiconductor solid sphere problem," *Alexandria Eng J*, vol. 127, pp. 1126–1142, Aug. 2025, doi: 10.1016/j.aej.2025.07.012.
[36] A. R. Keivanimehr and M. Akbari, "TinyML and edge intelligence applications in cardiovascular disease: A survey," *Comput Biol Med*, vol. 186, p. 109653, Mar. 2025, doi: 10.1016/j.compbiomed.2025.109653.
[37] D. Kumar, S. Puthusserypady, H. Dominguez, K. Sharma, and J. E. Bardram, "CACHET-CADB: A Contextualized Ambulatory Electrocardiography Arrhythmia Dataset," *Front Cardiovasc Med*, vol. 9, Jul. 2022, doi: 10.3389/fcvm.2022.893090.
[38] M. Meier, B. U. Demirel, and C. Holz, "WildPPG: A Real-World PPG Dataset of Long Continuous Recordings," Dec. 2024, doi: 10.3389/fcvm.2022.893981.
[39] S. Park, D. Zheng, and U. Lee, "A PPG Signal Dataset Collected in Semi-Naturalistic Settings Using Galaxy Watch," *Sci Data*, vol. 12, no. 1, p. 892, May 2025, doi: 10.1038/s41597-025-05152-z.
[40] D. Han *et al.*, "Multiclass arrhythmia classification using multimodal smartwatch photoplethysmography signals collected in real-life settings," December 13, 2024. doi: 10.21203/rs.3.rs-5463126/v1.
[41] M. Hosseini *et al.*, "A multimodal stress detection dataset with facial expressions and physiological signals," *Sci Data*, vol. 12, no. 1, p. 1844, Nov. 2025, doi: 10.1038/s41597-025-05812-0.
[42] A. N. Burkitt, "A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input.," *Biol Cybern*, vol. 95, no. 1, pp. 1–19, Jul. 2006, doi: 10.1007/s00422-006-0068-6.

BIOGRAPHIES OF AUTHORS

Anirudha Gaikwad is an Assistant Professor, Software Developer, and Freelance Corporate Technical Trainer with over 17+ years of industry experience. Renowned for his passion for software development and education, he blends innovation with practical insight to deliver impactful learning experiences. Anirudha has mentored over 5000 professionals through instructor-led training, academic project guidance, and corporate development programs. His hands-on approach bridges the gap between theory and real-world application, empowering learners to excel in modern tech environments. He has delivered more than 40 software projects in

the web and Android domains. Anirudha excelled in project management, client interaction, and cross-disciplinary collaboration. His dedication to innovation and education continues to drive success in both academia and industry. Anirudha's commitment to skill development, client collaboration, and continuous innovation positions him as a transformative figure in the tech training ecosystem. He can be contacted at email: gaikwadanirudha@rediffmail.com

Atit Gaikwad is a skilled educator and electronics professional with over nine years of experience in product design, circuit development, and system validation. Currently a Lecturer in the Electronics and Telecommunication Department at S.P.M. Polytechnic, Solapur, he combines academic expertise with industry knowledge to prepare students for practical challenges. He can be contacted at email: atitgaikwad@gmail.com

Dheeraj Lokhande has a Ph.D. in computer science and Engineering. He has 17 years of teaching experience. He has published 23 research papers. He has written 4 books. He has published 2 Patents. Also, he has qualified for the GATE EXAM. He can be contacted at email: dheerajlokhande1985@gmail.com