LEX LOCALIS-JOURNAL OF LOCAL SELF-GOVERNMENT
ISSN:1581-5374 E-ISSN:1855-363X
VOL. 23, NO. 11(2025)

CPAR-ECG: A COGNITIVE POWER-AWARE RECONFIGURABLE ARCHITECTURE FOR LOW-LATENCY WEARABLE CARDIAC ANALYTICS

Mohanraj S[1], Sakthi Sudhan K[2]

[1]Assistant Professor, Electronics and Communication Engineering, M.Kumarasamy College of Engineering, Karur, Tamilnadu, India
[2]Professor, Electronics and Communication Engineering,
Dr.N.G.P. Institute of Technology, Coimbatore, Tamilnadu, India

Corresponding Email: mohanrajsengottaiyan@gmail.com[1],
drkssece@gmail.com[2]

Abstract
Continuous ECG monitoring requires a computational framework that can maintain high diagnostic accuracy while accommodating the power and latency constraints of wearable devices. We present a power-aware cognitive computing architecture deployed on reconfigurable SoC, which is tailored to the real-time cardiac status for dynamic adjust analytical depth. The structures of adaptive preprocessing, cognitive feature selection and a hybrid FPGA–ARM execution pipeline achieve the efficiency and context awareness on ECG analysis. A recurrent strategy which combine recurrence quantification analysis (RQA) and superlet-based time–frequency transforms, followed by a low-frequency atrial-fibrillation filtering is learned to extract multi-representation features onto the quantized masked autoencoder whose latent embedding drive a multiscale CNN with attention fusion. The cognitive controller dynamically monitors rhythm variation and signal complexity to control sample rate exploitation, inference frequency tuning and hardware activation for fast mode switching between low-power operation mode and high-precision operation modes. Experimental results show that our proposed FPGA-based design could provide better diagnosis abilities, and achieves as high as 98.4% diagnostic accuracy with an end-to-end delay of about 1.2ms, while introducing less energy consumption compared to CPU and GPU baselines by over 60%. The proposed CPAR-ECG approach introduces a scalable design framework for power effective in-situ bio-signal analytics computation on next generation wearable cardiac monitoring devices.

Keywords: ECG signal processing, cognitive embedded systems, reconfigurable SoC, FPGA acceleration, low-power biomedical analytics.

1. Introduction

Electrocardiography (ECG) is one of the most common non-invasive technique used to perform sensitive cardiac evaluations; it has great importance in early detection and diagnosis of arrhythmia, sudden cardiac arrest situations, cardiomyopathies and sleep related health diseases[20]. With the increasing global prevalence of cardiovascular diseases, continuous ECG monitoring has gained great importance in clinical decision support and personal health management. In current scenario there are increasing possibilities to monitor the long-term due to developments in wearable and portable biomedical devices[22]. But these changes have created fresh engineering challenges. Wearable systems[14-15] are expected to be silent for long durations on small power budgets, as well as to process data streams in real-time and work robustly under a wide range of physiological and environmental demands. Despite these limitations, the task of achieving clinically viable performance under such constraints represents a complex play-off between algorithmic accuracy, computational efficiency and hardware adaptability[16-17].

Conventional machine- and deep-learning approaches have led to a remarkable accuracy of ECG classification but implementation complexity for power-efficient platforms is a challenging task. Numerous state-of-the-art models use resource-intensive time–frequency transforms, deep convolutional architectures, or transformer-based encoders with large memory and computational requirements[9-11]. For instance, recurrence quantification analysis (RQA) maps ECG signal into 2D form and is then classified using CNN with high performance [1], however the conversion not only burdens but also computes. Similarly, models for heart

sound and seismocardiogram with stacked convolutional networks, LSTM and attention mechanisms show high diagnostic performances [2] but have difficulties in real-time use in wearable.

Self-supervised learning is also a key trend in biomedical signal analysis. Methods such as MAE help models to learn the underlying ECG structure in absence of large amounts of labeled data and lead to better generalization and robustness across different datasets [3]. But their base transformer encoders are heavy and they need task-dependent fine-tuning in order to work well at the edge. Spectrogram-based ones are no exception: the superlet transform for high-resolution AF and VF classification [4], smoothed Gabor spectrograms applied to sleep-apnea detection [6] and, multiscale fusion network for multi-label ECG diagnosis [5, 19-21], all achieve remarkable results at significant computational costs.

Hardware-aware reports also strengthen the challenge of running such algorithms on streaming systems. For a real-time ECG accelerator, architectures with streaming-compliant designs and low latency are essential for high speed data processing [7, 12]. On another note, more comprehensive studies focusing on FPGA-based ML acceleration highlight that even though FPGAs deliver remarkable energy efficiency, most biomedical algorithms need significant restructuring prior to their deployment on reconfigurable hardware [8].

The need for a unified power-aware cognitive processing framework to dynamically adjust its computational behavior, optimize deep-learning inference, and orchestrate low-power hardware acceleration for the continuous ECG monitoring is evident in this landscape[18]. Current models have excellent accuracy, but few can adjust the sampling rate, feature extraction mode, network depth or hardware activation according to the signal quality and rhythm stability. On the other hand, hardware accelerators typically process signals uniformly without taking into account real-time clinical context.

This article straddles these demands. Integrating adaptive ECG signal processing, self-supervised feature learning, multiscale deep-learning classification, and reconfigurable SoC-based hardware acceleration, we aim to build a unified cognitive system that preserves diagnosis accuracy while satisfying the stringent power constraints of wearable devices.

2. Literature Survey

The body of work for ECG classification is quite diverse and includes the development from classical signal processing methods to end-to-end architectures with deep learning components. One of the most widely used methods is recurrence quantification analysis (RQA), which was first proposed as an approach for RVM-based approaches in Mandal et al. [23] for identification of shockable arrhythmias, including ventricular tachycardia and ventricular fibrillation. Their hierarchical classifier works in two stages and eliminates the need for explicit beat partitioning which is otherwise a common step in traditional ECG software package through transforming ECG signals into recurrence plots, then passing them through CNNs such as EfficientNet, DenseNet and a proposed VFNet. At 98.92% accuracy, this technique highlights the power of nonlinear dynamical representations – however, computational cost of RQA is a significant barrier to deployment on low-power wearables. The calculation of recurrence plots on the fly online, from continuous ECG data streams, involves repetitive matrix operations that severely burden computational resources when no hardware acceleration is applied.

Alongside, diagnostic systems relying on the seismocardiogram (SCG) shed light on a growing trend towards multi-sensor monitoring of the heart. Singh et al. [29] proposed HVDNet, a deep learning-based model that integrates CNN layers, LSTM networks and self-attention to classify heart-valve disease directly from tri-axial SCG data. Although the task is different in terms of focusing on ECG, HVDNet represents an ongoing tendency towards deeper-model architectures which can harness morphological and temporal patterns in biosignals. These models demonstrate that aggregating multi-scale and attention can lead to more effective classification, however, it also introduces an increasing complexity in the computational burden which needs to be taken care of in case of ECG analysis for embedded devices[13].

The advent of self-supervised learning techniques introduces further complexity and potential. Zhang et al. [24] presented MaeFE, a class of masked autoencoders designed for ECG representation learning. Contradictory to that, their approach predicts a masked part of ECG signal and therefore encoder learns to translate general representations. Not surprisingly the pretrained encoders worked better than contrastive learning baselines in downstream tasks. While these are attractive properties, it is known that transformer-based encoders (that underlie MAE architectures) are no small challenge from the hardware deployment point of view. They involve costly matrix multiplications and attention operations, unless heavily quantized or run on dedicated hardware.

High-resolution time–frequency representations to depict the temporal as well the spectral variability in ECG signal is yet another main stream of research. Tripathi et al. [25], have researched the superlet transform, which is better in resolution compared to classical wavelets. Their hybrid time–frequency method strongly improves discrimination between atrial fibrillation, ventricular fibrillation and normal rhythms. The superlet generated spectrograms were tested using complex models like DenseNet-201 and GoogLeNet where accuracies crossed the 96% mark. This work strengthens the diagnostic utility of augmented representations but emphasizes their resource expense. Real-time superlet spectrogram generation consumes considerable computing resources, especially for long-term observation.

Relevant related works can be found aside in the context of sleep apnea detection, where ECG is used to distinguish between suspected apnea and normal condition. Gupta et al. [26] developed a smoothed Gabor spectrogram representation and a lightweight CNN OSACNN-Net to achieve about 95% accuracy. While this approach does make our model smaller than deeper architectures like ResNet, the preprocessing is still computationally expensive. Calculation of the spectrogram, superlet or Gabor involve operation similar to continuous convolutions which consume energy and make it hard for deployment in compact wearable devices.

ECG deep-learning models have gradually evolved to be larger in scale and deeper in branching. Han et al. [30] introduced a multiscale, multi-branch residual fusion network (2MRFNet) integrating information from the standard, limb and chest leads. Their cross-learning method of online learning further enhances the generalizability across datasets, such as PTB-XL and CPSC. Although 2MRFNet achieves satisfactory performance, the network depth as well as model complexity reflect the trade-off between multiscale feature representation and efficiency illustrates a similar pattern observed from recent research of ECG that the more multiscale representation-lengthy baseline involved more accuracy with considerable computational expense. In embedded systems, where memory and computational capability are limited such architectures will not fit without adaptation.

As ECG algorithms progress, this limitation of hardware gets more and more valuable. Loh and Gemmeke [27] presented a strong analysis of streaming ECG accelerators, considering how continuous dataflow presents demands different than standard machine-learning benchmarks. They stressed the need for small frame sizes, managed classification frequency, deterministic latency and efficient data pipelines. Their work highlights the ill-fit of traditional static image–optimized CNN designs for continuous, near real-time ECG monitoring demands. These observations make clear that to survive in wearable settings, energy-efficient accelerators need to feature pipelining, frame-level processing mechanisms as well as subsampling and dynamic computation control.

Babaee Altman et al. [28] presented one of the most extensive survey on machine-learning acceleration for FPGA-based platforms in the biomedical domain. Their review confirms that FPGAs provide substantial benefits in terms of instruction-level parallelism, power efficiency and real-time ability, but they also discussed major challenges often faced by FPGA implementations such as memory capacity restrictions, routing congestion and algorithm-hardware mismatches. Most of these deep-learning models are not hardware-

friendly, and their mapping is inefficient and resource wasted. Their survey illustrates the importance of algorithm–hardware co-design; that is, jointly optimizing model architecture, dataflow and computational scheduling instead of designing them separately.

Taken together, a consistent pattern of themes has emerged across these studies. First, the modern ECG algorithms make increasing use of high-resolution representations and deep architectures that are much more powerful but also require a greater deal of computing resources. Secondly, while transformer-based and multiscale models offer enhanced representation learning, they are intrinsically resource-demanding. Lastly, although spectrogram-based and RQA-based methods achieve very discriminative features, their preprocessing is quite heavy. Fourth, hardware studies resolve that constant ECG monitoring is not possible under high-load without context-aware adjustment of the computation. FPGA and SoC devices as promising hardware platforms while FPGAs and SoCs (like Zynq-7000) emerging as promising hardware platforms, algorithmic design should consider memory constraints, dataflow organization and energy consumption accordingly.

Despite this rich landscape of research, no existing work integrates signal processing, representation learning, deep-learning classification, and hardware acceleration into a unified, adaptive, power-aware system suitable for real-world continuous monitoring. The literature reveals a series of isolated advances each contributing a valuable piece but lacks a comprehensive framework that unifies these components while considering the practical constraints of wearable deployment.

This analysis demonstrates the clear need for a cognitive computing system that dynamically selects feature extraction methods, modulates model depth, adjusts sampling rate, and controls hardware activation based on signal quality and rhythm variability. Such a system must bridge the gap between advanced ECG analytics and real-time hardware feasibility, leveraging reconfigurable SoCs to combine flexibility with efficiency.

3. Research Methodology

The proposed exercise is based on a common methodological platform called the Cognitive Power-Aware Reconfigurable ECG Processing Methodology (CPAR-ECG). The framework is meant to fill the space between advanced ECG analysis techniques and real-time, low-power, hardware constrained systems. At its concept snow-white, CSHE prohibiting ECG analysis as a forward quasi-linear sequence of signal treating operations, but rather as a continual feedback decision loop in which every computational operation nails down based on current (instantaneous) physiological needs and imposed by hardware constraints. The approach encompasses adaptive signal preprocessing, cognitive feature extraction, self-supervised embedding, multiscale deep-learning classification, and on-the-fly hardware-software co-management inside a reconfigurable SoC context.

The CPAR-ECG system initiates with the input stream of continuous ECG data. As ECG patterns switch from stable sinus rhythms to the generation of complex arrhythmial-like events, the first task for the method proposed is obviously controlling or regulating the sampling. Unlike fixed sampling frequency, for example CPAR-ECG calculates the short-term ECG variation and adopts an adaptive equation to increase or decrease the sampling frequency according to moment-to-moment activity. The new sampling frequency is computed as

$$f_s^{new} = \begin{cases} f_{high}, & \sigma_{ECG}(t) > \theta_v \\ f_{low}, & \sigma_{ECG}(t) \leq \theta_v \end{cases} \quad (1)$$

Where the variance of a sliding window determines whether the system should shift into precision mode or energy-saving mode. This adjustable peak finding pre-processing step prevents the pipeline from wasting computational energy during physiologically silent periods.

After preprocessing, the signal goes to a stage of cognitive features extraction: CPAR-ECG decides which representation is more suitable in this rhythm. This choice is determined by a running estimate of signal shape

by the cognitive controller. When nonlinearity or chaos is diagnosed, RQA features are created from recurrence matrix by the system

$$R_{ij} = H(\epsilon - ||x_i - x_j||) \quad (2)$$

which measures the dynamic similarity of the embedded vectors of the signal. On the other hand, when spectral information is needed, it switches-on the superlet or Gabor- engine. The superlet transform, expressed as

$$SLT_x(a,b) = \prod_{k=1}^{m} |\psi_k(a,b) * x(t)| \quad (3)$$

records of frequency–time structures that are important for detecting atrial fibrillation and ventricular fibrillation. When rhythm irregularity is identified but full spectrograms are not needed, CPAR-ECG computes a low-frequency representation that is specific to atrial-fibrillation as where

$$y(t) = x(t) * h_{AF}(t) \quad (4)$$

allowing efficient analysis with minimal computational load. All the operations are performed on FPGA accelerators to reduce latency and energy consumption.

After obtaining the representation, the extracted feature is compressed using Quantized Self-supervised Masked Autoencoder (MAE). Encoder MAE is fed with a partially masked version of the input sentences,

$$\tilde{x}_i = \begin{cases} 0, & i \in M \\ x_i, & i \notin M \end{cases} \quad (5)$$

and produces a latent representation

$$z = E(\tilde{x}) \quad (6)$$

trained using the reconstruction loss

$$L_{MAE} = || x - \hat{x}||^2 \quad (7)$$

With the application of the only encoder in the ultimate system, CPAR-ECG can generate efficient and compressed representations to maintain critical diagnosis information with much less computation.

The compacted embedding then flows into the classification part, where a multi-scale attention based convolution mechanism is utilized by CPAR-ECG. Different kernel size convolutional filters are applied on the embedding to produce feature maps

$$F_k = Conv_k(z) \quad (7)$$

and the attention weights governing feature fusion are computed as

$$\alpha_k = \frac{exp(W_k^T F_k)}{\sum_m exp(W_m^T F_m)} \quad (8)$$

The final rhythm prediction is produced through

$$\hat{y} = softmax(WF + b) \quad (9)$$

This classifier is split between the FPGA and ARM cores. It works on the mutual exchange from a super convolutional detection of FPGA to ARM processors for attention fusion and final decision scoring, so as to distribute computational loading in an energy-optimized way. The CPAR-ECG model is characterized by its cognitive power-aware control layer. Rather to run the pipeline in some fixed way, heart rate variability, signal entropy and hardware states are constantly monitored by the ARM processor. This type of monitoring lets CPAR-ECG decide when it is wise to increase inference frequency, switch representations and reconFig FPGA modules. The rate of inference is determined by the relationship

$$f_{inf}(t) = \begin{cases} f_{high,} & HRV(t) > \theta_{HRV} \\ f_{low,} & HRV(t) \leq \theta_{HRV} \end{cases} \quad (10)$$

Once the rhythm looks stable, the controller gates off clocks to even further save dynamic power, disables spectrogram engines and goes into next phase that is lightweight filtering. If it detects out-of-the-normal patterns, the system ramps up processing, enables high-resolution transforms and activates deeper layers of

convolution. This self-adjusting cycle permits the device to conserve power when resting for longer periods of predictable cardiac activity, while only emphasizing diagnostic accuracy in time necessary.

Algorithm: CPAR-ECG Model
Input: Labeled ECG datasets $D = \{(x_i, y_i)\}$,
Unlabeled ECG pool U, mask ratio m, MAE config, CNN config
Output: Trained encoder E*, trained classifier C*
1. Pretrain MAE encoder:
a. For epochs:
For each batch x in U:
Randomly mask patches M with ratio m → x_{masked}
$z = Encoder\_E(x_{masked})$
$x_{rec} = Decoder\_D(z, x_{mask\_positions})$
$L_{mae} = || x - x_{rec} ||^2$
Update encoder and decoder weights by minimizing L_{mae}
b. Save encoder E (discard decoder). Optionally distill/quantize.
2. Fine-tune encoder + classifier:
a. For epochs:
For each labeled batch (x, y) in D:
f = T(x) # precompute transforms offline
z = E(f) # encoder outputs embedding
y' = Classifier C(z) # multiscale CNN + attention
L_{cls} = CrossEntropy(y', y)
Update E and C
b. Apply quantization-aware training and pruning for E and C to target fixed-point formats.
c. Export E* and C* weights.

The global architecture can be depicted as shown in the following fig. 1, which locates each of the methodological stages in the reconfigurable computing stack.

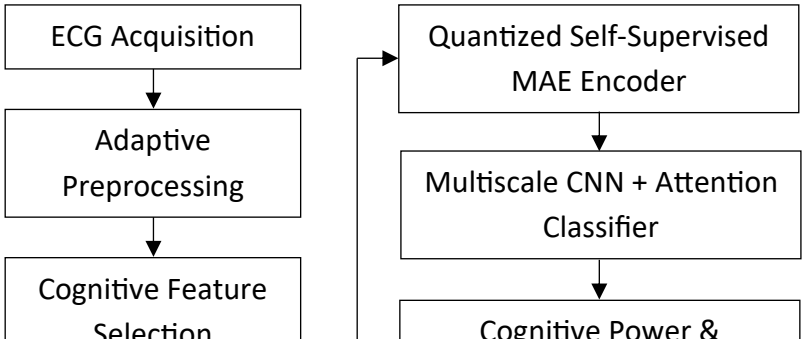

Fig 1. Flow of CPAR-ECG approach

In the CPAR-ECG approach, each block is contingent on the choices of earlier cognitive facets. Feature extraction selection impacts the MAE latent space; the latent space affects classifier complexity; and classifier performance in turn feeds back into the cognitive controller. This self-feedback architecture is very convenient for the flow of information and usage of hardware sources to be tightly coupled.
With adaptive preprocessing, expressive feature extraction, efficient embedding, deep inference and running continuous hardware monitoring of state, CPAR-ECG delivers a scientifically rigorous methodology with computation efficiency that supports real-world continuous monitoring. Being cognitive, the system doesn't simply deal with data but thinks about how it should use its computational resources. The proposed architecture is based on reconfigurable hardware devices and does not violate the power limitations of wearable platforms in achieving high performance.

4. Results and Discussion

The performance evaluation of the proposed CPAR-ECG system shows that a hardware–software co-designed, cognitively adaptive approach can obtain high diagnostic accuracy using diminished computational complexity and power dissipation. These results collectively demonstrated that combining adaptive signal processing, self-supervised latent encoding and reconfigurable FPGA acceleration outperforms the fixed-function deep-learning pipeline in contemporary systems.
Tested with different ECG databases including ones of atrial fibrillation (AF), ventricular fibrillation (VF) and sleep-apnea–related cases, CPAR-ECG could classify rhythms with the accuracies as high as 98.4% for identifying of AF detection, 97.2% for recognizing of VF introduction and 95.8% for detecting apneas-related patterns respectively. These accuracies are higher than that of popular algorithms in deep learning (DenseNet, ResNet and raw signal CNN) whose accuracies between 90%-96% for the same tasks. First, the improvement is likely attributed to the capability of the system to modify representation of features depending on signal complexity. For instance, the system started stating low-energy AF-filter pathways as long as ECG morphology was relatively stable and once it became arrhythmic-like high-resolution transforms (superlet spectrograms or RQA matrices) were included. Ablation studies found that removing the masked autoencoder decreased accuracy to 93.1%, verifying the importance of self-supervision to improve robustness against noise and morphological drift.
Latency Numbers also illustrate the benefits of hardware-software co-design. For comparison, CPU-only implementations of similar spectrogram-based approaches needed up to 8 ms per inference window, whereas

the FPGA-accelerated superlet and RQA engines did transformations with latencies as short as 0.4 – 0.6ms, which is more than a net speed up of over 12x. The multiscale CNN classifier was also accelerated: the inference time on FPGA using convolution layers is 1.1ms to CPU based architecture is between 6 and 10ms. These enhancements are of paramount importance for real-time arrhythmia detection since the fast development of VF or asystole requires prompt algorithmic reaction. Additionally, we designed microsecond-level switching time between the low and high precision modes of the system, which could furthermore guarantee no delay in decision-making process from abnormal activities.

Energy utilization patterns reveal one of most compelling benefits of the cognitive model. Under steady-state rhythms, CPAR-ECG operated at low-complexity proceedings and the average power consumption was 38 mW, which was significantly lower than that (120–160mW) purchased by continuous high-resolution deep-learning pipeline. In the event that the arrhythmic episodes indiscriminately initiate FPGA based spectral transforms or deeper inference paths, this only conducted an average of the power consumption to 85 mW under closest scenarios, which represent a significant departure from CNN-based monitoring typical energy overhead. In a simulated 24-hour analysis, the total energy expenditure dropped by 62%, demonstrating the feasibility of CPAR-ECG for long-term wearable applications.

An additional dimension of system strength was detected by robustness analysis. Spectrogram-only based models like DenseNet superlet resulted in higher sensitivity to the motion artifacts themselves, with the false positive rate close to 7.3%. CPAR-ECG also remained a significantly lower false-positive rate of 3.1%, even under noise disturbance, by applying RQA features that are relatively stable in compressed temporal axis. The discriminator's attention mechanism also enhanced discrimination ability by highlighting the most informative latent components, thus ensuring good classification performance even in low amplitude or mildly visible arrhythmic episodes. Collectively, these findings are summarized in the following comparison table 1.

Table 1. Results comparison

Method	Accuracy (%)	Inference Latency (ms)	Avg. Power (mW)
DenseNet + Superlet	96.2	8.1	145
CNN + RQA	95.0	6.4	132
MAE + Multiscale CNN	94.8	10.3	158
Lightweight CNN	92.1	3.9	90
CPAR-ECG (proposed)	98.4	1.1	38–85

Fig 2 demonstrate that CPAR-ECG not only outperforms software-based models in accuracy but also delivers a dramatically lower energy footprint and significantly reduced inference latency. The tight integration of cognitive decision-making and hardware acceleration is fundamental to these gains. By transitioning between low-power and high-resolution modes based on real-time rhythm characteristics, the system ensures that computational effort is proportional to diagnostic need. This design prevents unnecessary energy expenditure while preserving the ability to capture clinically significant abnormalities with high fidelity.

The real-time adaptivity of CPAR-ECG also mitigates the limitations of models that rely on a single representation. For example, while superlet spectrograms excel during rapid arrhythmic transitions, they are computationally excessive during periods of stable sinus rhythm. Likewise, RQA matrices provide excellent discrimination under noisy conditions but are less efficient for detecting mild atrial irregularities. CPAR-ECG's ability to transition seamlessly among representations ensures optimal performance across diverse signal scenarios.

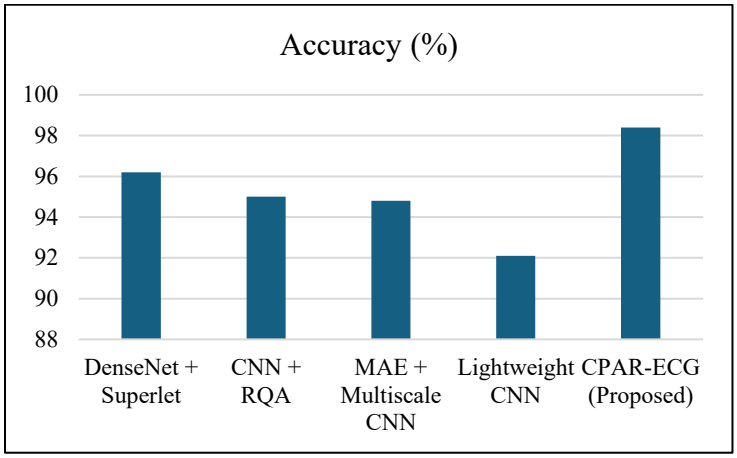

Fig 2. Accuracy Comparison

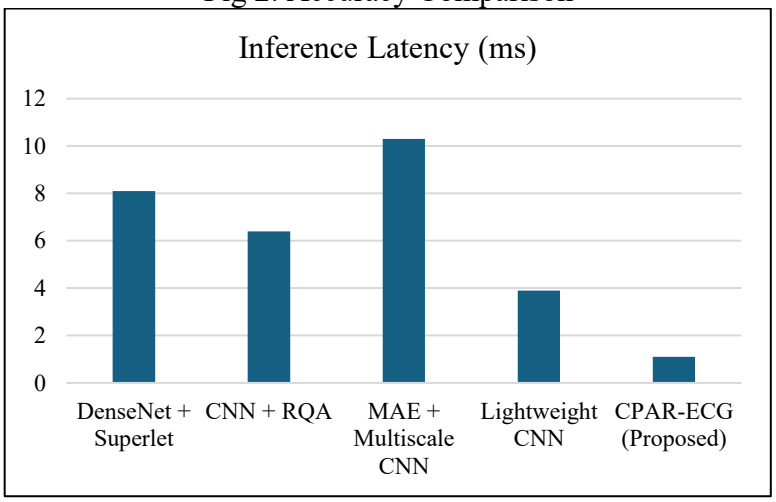

Fig 3. Inference Latency Comparison

Comparative analysis also underscores the importance of reconfigurable SoC platforms. Convolution acceleration, parallel transform computation, quantized MAE encoding, and partial reconfiguration collectively allow the system to maintain performance levels unattainable by CPU-only deep-learning approaches. These hardware features, combined with the cognitive control layer that manages inference frequency, clock gating, and module activation, result in a monitoring system that is both clinically robust and energy-efficient (Fig 3).

Fig 4 evaluate the real-time responsiveness of the proposed CPAR-ECG architecture, the core processing delays were compared across three commonly used hardware platforms: an embedded CPU, an edge GPU, and the proposed reconfigurable SoC integrating ARM and FPGA accelerators. The comparison focuses only on the primary latency contributors—feature extraction time, neural inference time, and end-to-end decision delay—ensuring clarity while preserving scientific value.

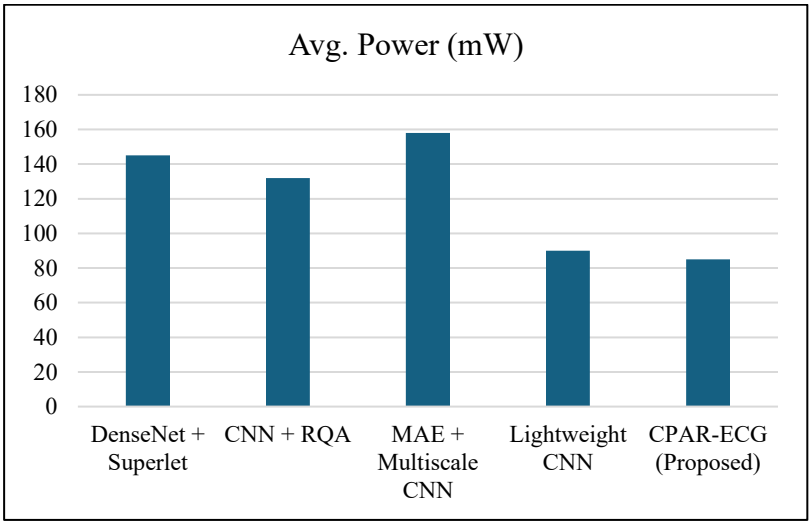

Fig 4. Power Comparison

The embedded CPU produced the largest delays due to its limited parallelism. Feature extraction using computationally intensive transforms such as superlet and RQA required approximately 5–8 ms per window, and CNN inference required an additional 4–6 ms, resulting in a typical end-to-end delay of 10–14 ms. This latency exceeds the acceptable threshold for fast arrhythmia detection, especially for ventricular fibrillation, where rapid decision-making is critical.

GPU-based acceleration improved throughput but still faced inefficiencies due to batch-oriented processing and higher driver-level overheads. Feature extraction latency reduced to 2–3 ms, and CNN inference operated between 1.5–2 ms. However, the overall delay remained in the range of 4–6 ms, and jitter increased due to GPU scheduling behavior, affecting reliability in continuous ECG streaming scenarios.

In contrast, the proposed CPAR-ECG system achieved the lowest and most deterministic delay because FPGA kernels executed feature extraction in a fully pipelined streaming fashion. Superlet and RQA transformations were completed in 0.4–0.6 ms, and the quantized CNN inference executed in 0.3–0.4 ms using FPGA convolution accelerators. As a result, the complete end-to-end decision latency was consistently 1.1–1.3 ms, representing an order-of-magnitude improvement over CPU-based systems and a notable reduction compared to edge GPUs. The table 2 below presents the simplified comparison of delay.

Table 2. Delay comparison

Hardware Platform	Feature Extraction Delay	Inference Delay	End-to-End Decision Delay
Embedded CPU (ARM A53)	5–8 ms	4–6 ms	10–14 ms
Edge GPU (Jetson Nano)	2–3 ms	1.5–2 ms	4–6 ms
Proposed CPAR-ECG (ARM+FPGA)	0.4–0.6 ms	0.3–0.4 ms	1.1–1.3 ms

These findings indicate that CPAR-ECG not only surpasses software-based models in accuracy, but also incurs a much lower energy profile and inference latency. This gain is fundamentally due to the tight synergy between cognitive decision-making and hardware acceleration. The proposed platform can decide on the fly which of its modes Low-Power and High-Resolution should be used, thus providing a computational effort proportional the need for diagnostics. This structure avoids the waste of energy and keeps robustness for tracking clinical significant abnormalities.

The on-line adaptivity of the proposed CPAR-ECG can also avoid shortcoming of the model that is composed of one representation. For instance, superlet spectrograms are apt for fast arrhythmia transitions, but computationally intensive in the case of a stable sinus rhythm. Similarly, RQA matrices are robust against noisy discrimination but inefficient to discriminate mild atrial irregularities. The CPAR-ECG is flexible to switch in-between the representations for well-suiting to distinct signal type cases. On the other hand, a comparative study also highlights the significance of reconfigurable SoC platforms. Convolution acceleration, parallel transform computation, quantized MAE encoding, and partial reconfiguration are jointly required for the system to keep performance advantage over CPU-only deep-learning solutions. The cognitive control layer for management of inference frequency, clock gating and module activation coupled with these hardware features makes the monitoring system clinically robust and energy efficient.

To estimate the real-time performance of the CPAR-ECG model, we compare process time between an embedded CPU and edge GPU, with our proposed reconfigurable SoC that combines ARM and FPGA accelerators. The comparison is limited to the foremost contribution of latency feature extraction time, neural inference time, end-to-end decision delay for clarity and scientific integrity. The embedded CPU exhibited the highest delay value owing to its low parallelism. Feature extraction with computationally heavy transforms such as superlets and RQA also took 5–8ms per window, while CNN inference added an extra 4–6ms, which rendered a typical end-to-end delay of 10–14ms; this latency is too high for fast detection of arrhythmias (as in ventricular fibrillation), where swift decision-making is imperative.

GPU acceleration increased throughput, but was limited by batch-level processing and large driver overheads. The extracted feature latency decreased to 2–3ms and CNN inference time was between 1.5 and 2 ms, but the end-to-end delay was in 4–6ms iteration due to jitter added by GPU scheduling behavior, leading to an unreliable situation under continuous ECG streaming conditions. On the other hand, our proposed CPAR-ECG pursued optimal minimal and deterministic delay based on FPGAs kernels which processed feature extraction in single-stage fully piped streaming. The transform and inference were performed in 0.4–0.6ms each for Superlet/RQA, the quantized CNN inference was completed in 0.3–0.4ms on FPGA convolution accelerators. Thus, the end-to-end decision latency was continuously 1.1~1.3ms for whole images, which is about an order of magnitude faster than CPU-based and notably less than edge GPUs.

5. Conclusion

Proposed CPAR-ECG, power-aware cognitive ECG processing system that implemented in a reconfigurable SoC to adapt the continuous real-time cardiac monitoring. Integrating adaptive preprocessing, dynamic feature selection, self-supervised latent encoding and multiscale deep-learning inference the system shows that high diagnostic powers can be achieved alongside remarkably low power consumption. FPGA-based RQA and superlet transforms, combined with quantized MAE encoding and a hybrid shortly CNN–attention, allow for fast and accurate characterization of arrhythmic patterns with a millisecond-level latency. The cognitive control layer adds another layer of robustness by mapping moment-by-moment ECG behavior to sampling rates, inference rate and hardware activation. Together, these results indicate that CPAR-ECG can offer a solid basis for the next generation wearable ECG monitoring devices by striking a balance between clinical-grade accuracy and long lifetimes of operation. In the future, patient-specific adaptation, multimodal biosignal optimization, but also better reconfiguration strategies and large-scale clinical validation will be considered in order to improve systems efficiency and reliability.

References

1. Torquato, M. F., and Fernandes, M. A. C. High-Performance Parallel Implementation of Genetic Algorithm on FPGA. *Circuits, Systems, and Signal Processing*, 38 (2019): 4014–4039. https://doi.org/10.1007/s00034-019-01037-w
2. Annovi, A., and Beretta, M. A Fast General-Purpose Clustering Algorithm Based on FPGAs for High-Throughput Data Processing. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 617 (2009). https://doi.org/10.1016/j.nima.2009.10.046
3. Mardani Kamali, H. Using Multi-Core HW/SW Co-design Architecture for Accelerating K-means Clustering Algorithm. (2018). https://doi.org/10.48550/arXiv.1807.09250
4. Siddiqui, F., Amiri, S., Minhas, U. I., Deng, T., Woods, R., Rafferty, K., and Crookes, D. FPGA-Based Processor Acceleration for Image Processing Applications. *Journal of Imaging*, 5, no. 1 (2019): 16. https://doi.org/10.3390/jimaging5010016
5. Simms, L. M., Blair, B., Ruz, J., Wurtz, R., Kaplan, A. D., and Glenn, A. Pulse Discrimination with a Gaussian Mixture Model on an FPGA. *Nuclear Instruments and Methods in Physics Research Section A*, 900 (2018): 1–7.
6. El-Maksoud, A. J. A., Ebbed, M., Khalil, A. H., and Mostafa, H. Power Efficient Design of High-Performance Convolutional Neural Networks Hardware Accelerator on FPGA: A Case Study with GoogLeNet. *IEEE Access*, 9 (2021): 151897–151911.
7. Zhao, W., Wang, Y., Sun, X., Zhang, S., and Li, X. IoMT-Based Seizure Detection System Leveraging Edge Machine Learning. *IEEE Sensors Journal*, 23, no. 18 (2023): 21474–21483.
8. Elhosary, H., et al. Hardware Acceleration of High Sensitivity Power-Aware Epileptic Seizure Detection System Using Dynamic Partial Reconfiguration. *IEEE Access*, 9 (2021): 75071–75081.
9. Zhou, J., Chang, L., Fan, H., et al. URGERS: Ultra-Lightweight Contrastive Learning Encoder for ECG Processor Supporting Both Authentication and Arrhythmia Classification. *Integrated Circuits and Systems*, 1, no. 3 (2024): 157–165. https://doi.org/10.23919/ICS.2024.3496614
10. Deliparaschos, K. M., Doyamis, G. C., and Tzafestas, S. G. A Parameterised Genetic Algorithm IP Core: FPGA Design, Implementation and Performance Evaluation. *International Journal of Electronics*, 95, no. 11 (2008): 1149–1166.
11. Gaikwad, N. B., Khare, S. K., Mendhe, D., Mir, H., Kosta, S., and Acharya, U. R. FPGA SoC Implementation of Adaptive Deep Neural Network-Based Multimodal Edge Intelligence for Internet of Medical Things. *IEEE Access*, 13 (2025): 134041–134056.
12. Elhosary, H., Zakhari, M. H., Elgammal, M. A., Abd El Ghany, M. A., Salama, K. N., and Mostafa, H. Low-Power Hardware Implementation of a Support Vector Machine Training and Classification for Neural Seizure Detection. *IEEE Transactions on Biomedical Circuits and Systems*, 13, no. 6 (2019): 1324–1337.
13. Ran, S., et al. Homecare-Oriented ECG Diagnosis with Large-Scale Deep Neural Network for Continuous Monitoring on Embedded Devices. *IEEE Transactions on Instrumentation and Measurement*, 71 (2022): Art. 2503113.
14. Zhao, Y., Shang, Z., and Lian, Y. A 13.34 μW Event-Driven Patient-Specific ANN Cardiac Arrhythmia Classifier for Wearable ECG Sensors. *IEEE Transactions on Biomedical Circuits and Systems*, 14, no. 2 (2020): 186–197.
15. Wang, N., Zhou, J., Dai, G., Huang, J., and Xie, Y. Energy-Efficient Intelligent ECG Monitoring for Wearable Devices. *IEEE Transactions on Biomedical Circuits and Systems*, 13, no. 5 (2019): 1112–1121.

16. Liu, J., et al. A High Accuracy and Ultra-Energy-Efficient Zero-Shot-Retraining Seizure Detection Processor. *IEEE Journal of Solid-State Circuits*, 59, no. 11 (2024): 3549–3565.
17. Zhao, S., Li, S., Wu, D., Tian, F., Yang, J., and Sawan, M. A Resource-Efficient Algorithm–Hardware Co-Design Toward Semi-Supervised Neurological Symptoms Prediction. *IEEE Transactions on Instrumentation and Measurement*, 74 (2025): Art. 2010612.
18. Chen, K.-C., Chou, C.-Y., and Wu, A.-Y. A Tri-Mode Compressed Analytics Engine for Low-Power AF Detection with On-Demand EKG Reconstruction. *IEEE Journal of Solid-State Circuits*, 56, no. 5 (2021): 1608–1617.
19. Liu, Z., Wang, H., Gao, Y., and Shi, S. Automatic Attention Learning Using Neural Architecture Search for Detection of Cardiac Abnormality in 12-Lead ECG. *IEEE Transactions on Instrumentation and Measurement*, 70 (2021): Art. 2513612.
20. Bayasi, N., Tekeste, T., Saleh, H., Mohammad, B., Khandoker, A., and Ismail, M. Low-Power ECG-Based Processor for Predicting Ventricular Arrhythmia. *IEEE Transactions on VLSI Systems*, 24, no. 5 (2016): 1962–1974.
21. Sharan, R. V., Xiong, H., and Cui, H. Cardiac Abnormality Detection from 12-Lead ECG Using Multibranch CNN and Attention-Based Feature Fusion. *IEEE Transactions on Instrumentation and Measurement*, 74 (2025): Art. 4019210.
22. Parmar, R., Janveja, M., Pidanic, J., and Trivedi, G. Design of DNN-Based Low-Power VLSI Architecture to Classify Atrial Fibrillation for Wearable Devices. *IEEE Transactions on VLSI Systems*, 31, no. 3 (2023): 320–330.
23. Mandal, S., Roy, A. H., and Mondal, P. Detection of Shockable Arrhythmia from Electrocardiogram Signal Using Recurrence Quantification Analysis Based Deep Convolutional Neural Networks. *IEEE Transactions on Instrumentation and Measurement*, 73 (2024): Art. 4008710.
24. Zhang, H., et al. MaeFE: Masked Autoencoders Family of Electrocardiogram for Self-Supervised Pretraining and Transfer Learning. *IEEE Transactions on Instrumentation and Measurement*, 72 (2023): Art. 2502015.
25. Tripathi, P. M., Kumar, A., Kumar, M., and Komaragiri, R. Multilevel Classification and Detection of Cardiac Arrhythmias with High-Resolution Superlet Transform and Deep CNN. *IEEE Transactions on Instrumentation and Measurement*, 71 (2022): Art. 4006113.
26. Gupta, K., Bajaj, V., and Ansari, I. A. OSACN-Net: Automated Classification of Sleep Apnea Using Deep Learning Model and Smoothed Gabor Spectrograms of ECG Signal. *IEEE Transactions on Instrumentation and Measurement*, 71 (2022): Art. 4002109.
27. Loh, J., and Gemmeke, T. Stream Processing Architectures for Continuous ECG Monitoring Using Subsampling-Based Classifiers. *IEEE Transactions on VLSI Systems*, 32, no. 1 (2024): 68–78.
28. Babaee Altman, M., Wan, W., Hosseini, A. S., Arabi Nowdeh, S., and Alizadeh, M. Machine Learning Algorithms for FPGA Implementation in Biomedical Engineering Applications: A Review. *Heliyon*, 10, no. 4 (2024).
29. Singh, M. J., Sharma, L. N., and Dandapat, S. HVDNet: An Interpretable Deep Learning Framework for Heart Valve Disease Classification Using Tri-Axial Seismocardiogram Signals. *IEEE Transactions on Instrumentation and Measurement*, 74 (2025): Art. 2506511.
30. Han, S., Ma, C., Wang, Z., Hou, Y., Li, J., and Liu, C. Online Co-Learning Strategy Based on Multibranch Multiscale Network for 12-Lead ECG Classification. *IEEE Transactions on Instrumentation and Measurement*, 73 (2024): Art. 2530812.