

Triple-Factorized attention framework with dynamic adaptive windowing for premature contraction classification

Jibin Wang

wangjibin0503@126.com

Bengbu Medical College

Research Article

Keywords: Dynamic electrocardiogram recognition, Triple-Factorized attention, Adaptive contextual aggregation

Posted Date: August 18th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10311264/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Triple-Factorized attention framework with dynamic adaptive windowing for premature contraction classification

Jibin Wang^{a,b,*}

^a*School of Medical Imaging, Bengbu Medical University, Bengbu 233030, China*

^b*School of Mathematics, Tianjin University, Tianjin 300350, China*

Abstract

Premature atrial contraction (PAC) and premature ventricular contraction (PVC) are prevalent arrhythmias with significant clinical implications. However, their transient nature and subtle morphological variations challenge reliable detection by conventional electrocardiogram (ECG) analysis. To address this, we propose historical recognition with dynamic adaptive window network (HRDAWNet), a novel attention framework integrating Triple-Factorized Attention and Dynamic Adaptive Windowing for dynamic PAC and PVC recognition. Triple-Factorized Attention mechanism synergistically encodes channel-spatio-temporal dependencies to resolve the limitation of isolated feature processing in existing methods. Dynamic Adaptive Windowing also exploits learnable relevance scoring to selectively aggregate contextual embeddings from informative historical states to overcome the fixed-length context constraints. Validated on the China Physiological Signal Challenge 2018 (CPSC2018) and MIT-BIH arrhythmia (MITBA) databases, HRDAWNet achieves stable improvement with 99.3% accuracy (CPSC2018) and 99.0% accuracy (MITBA) compared to the existing methods. Ablation studies confirm the model’s robustness to physiological variability, rapid convergence and clinical interpretability via gradient-weighted class activation mapping, which localizes the critical features such as inverted P-waves in PAC and

*Corresponding author

Email address: wangjibin0503@126.com (Jibin Wang)

wide QRS complexes in PVC. These advancements establish a new paradigm for adaptive ECG analysis and demonstrate significant potential for real-time arrhythmia monitoring in clinical and wearable settings.

Keywords: Dynamic electrocardiogram recognition, Triple-Factorized attention, Adaptive contextual aggregation

1. Introduction

Cardiac arrhythmias represent a significant global health burden, implicated in over 32% of all deaths worldwide. While conduction abnormalities pose well-documented diagnostic challenges due to subtle electrocardiogram (ECG) variations and carry serious prognostic implications including heart failure and myocardial infarction, premature atrial contraction (PAC) and premature ventricular contraction (PVC) constitute another pervasive category of rhythm disturbances demanding urgent clinical attention. Though often initially perceived as benign, particularly in isolation, accumulating evidence underscores their profound clinical significance [1, 2]. Frequent PAC is strongly associated with an elevated risk of atrial fibrillation, stroke and increased mortality, while PVC can induce cardiomyopathy, heart failure and significantly increase the risk of sudden cardiac death. Critically, PAC and PVC are frequently asymptomatic or present with non-specific symptoms, leading to under-detection in routine clinical practice. Their transient and sporadic nature makes reliable identification on standard short-duration ECG challenging. Consequently, the development of accurate and automated methods for the timely detection of PAC and PVC is imperative. Such advancements are essential not only for risk stratification in seemingly healthy individuals and those with structural heart disease but also for guiding targeted therapeutic interventions to prevent severe cardiovascular complications.

While ECG remains the cornerstone for diagnosing arrhythmias, reliance on conventional visual interpretation presents significant limitations [3]. The labor-intensive nature of manual analysis, inherent susceptibility to human error and

increasing volume of ambulatory ECG recordings render traditional methods increasingly impractical and unsustainable in modern cardiology. These constraints are particularly critical for PAC and PVC detection, given their often subtle, transient morphological complexity and high prevalence. Consequently, overcoming the inherent subjectivity and diagnostic inconsistency of human-centric ECG review is paramount. This underscores an imperative clinical need: the development and deployment of robust, efficient computational approaches for the automated, accurate identification of PAC and PVC to enhance diagnostic reliability and scalability.

Despite considerable progress in automated ECG analysis, current methodologies predominantly rely on conventional machine learning (ML) algorithms requiring extensive handcrafted feature engineering [4]. This domain expertise-dependent process, utilizing morphological, statistical or transform-based techniques, significantly dictates algorithmic performance. However, the fundamental limitations impede the practical translation of these conventional ML approaches for robust PAC and PVC recognition. Challenges primarily stem from:

- 1) Limited generalizability, where models trained on specific datasets often fail to perform accurately on broader, real-world populations due to physiological variability and noise;
- 2) Pronounced susceptibility to overfitting, exacerbated by the high dimensionality of potential features relative to often limited annotated ECG data, leading to poor performance on unseen examples;
- 3) The inherent subjectivity and absence of universally applicable, standardized protocols for optimal feature selection and extraction.

These critical shortcomings collectively undermine reliability and restrict the real-world clinical deployment of traditional ML-based detection systems, necessitating more advanced, generalizable solutions.

To address these fundamental limitations, deep learning (DL) methodologies have emerged as a transformative paradigm in the computational ECG analysis. Unlike conventional ML, DL architectures inherently circumvent the need for manual feature engineering and explicit classifier selection, instead autonomously learning hierarchical representations directly from raw or minimally

processed signals [5]. This capability has catalyzed significant performance breakthroughs across diverse domains, demonstrating prominent efficacy in complex pattern recognition scenarios. Within ECG analytics, DL frameworks—especially convolutional neural networks (CNN) for spatial feature extraction, long short-term memory networks (LSTM) for modeling temporal dependencies and attention mechanisms for contextual focus—have established themselves as foundational architectures. Empirical evidence consistently demonstrates their superior accuracy in detecting subtle arrhythmic patterns compared to traditional approaches, while concurrently exhibiting enhanced model robustness and generalization capacity across heterogeneous datasets. These intrinsic superiorities position DL as a critical technological enabler for developing clinically viable, high-precision automated systems for PVC and PAC identification.

While DL methodologies (particularly CNN for spatial features and LSTM for temporal dynamics) have shown promising results in ECG analysis, several fundamental limitations persist: 1) **Isolated feature processing** treats spatial (sample episode), temporal (rhythmic) and channel-wise (sample point) data in silos, neglecting their synergistic interactions; 2) **Fixed-length historical context** forces static pattern recognition, disregarding the variable inter-sample dependencies characteristic of PAC/PVC [6, 7]. Consequently, even state-of-the-art DL models struggle to achieve clinically requisite reliability in real-world ambulatory monitoring scenarios where PAC/PVC exhibit high morphological variability and transient occurrence.

To transcend these barriers, we propose historical recognition with dynamic adaptive window network (HRDAWNNet), a novel architecture explicitly for dynamic PAC and PVC detection. Its core innovations address the gaps in the existing methodologies: 1) **Triple-Factorized Attention** synergizes Channel-Spatio-Temporal joint encoding, Inter-agent Attention and Mode Attention to holistically model interactions between sample episode features (spatial), rhythmic evolution (temporal) and sample point correlations (channel); 2) **Dynamic Adaptive Windowing** replaces fixed-length histories with learnable temporal relevance scoring, selectively aggregating contextual embeddings only from in-

formative prior states critical for capturing dynamic heartbeats. This framework establishes a novel paradigm where adaptive channel spatiotemporal-contextual integration supersedes rigid pattern matching, directly translating to superior
90 clinical utility. Ablation studies on the China Physical Signal Challenge 2018 (CPSC2018) and MIT-BIH arrhythmia (MITBA) databases confirm the effectiveness of the model.

The main contributions of this work include the following three points:

- **Dynamic Adaptive Modeling:** We propose HRDAWNet with learnable relevance scoring and adaptive thresholds to enable selective capture of transient PAC and PVC features, which overcomes the limitations of fixed-window approaches in existing methods.
- **Integrated Feature Learning:** By introducing Triple-Factorized Attention, we transform arrhythmia recognition from static to dynamic mod-
100 eling effectively integrating Spatio-Temporal-Channel dependencies to resolve variable inter-sample interactions.
- **Clinical Validation:** Extensive ablation studies on CPSC2018 and MITBA databases demonstrate consistent performance gains, highlighting the model’s potential for broader clinical applications and also establishing a novel paradigm for adaptive ECG analysis.

2. Related work

Existing PAC and PVC detection research falls into two paradigms: conventional ML algorithms and DL frameworks.

Conventional ML algorithms: Sangaiah et al. [8] developed an intelligent
110 learning method with the help of wavelet coefficient features and hidden Markov model (HMM) for PAC and PVC signal detection. Rahul et al. [9] put forward an RR interval-based method with statistical features and support vector machine (SVM)/random forest (RF) classifiers for PVC/PAC classification. Nahak et al. [10] proposed a method utilizing wavelet scattering transform (WST) features

and SVM classifier for fragment-level ECG arrhythmia classification. However, conventional ML approaches for automated PAC and PVC recognition suffer from the primary limitations: 1) poor generalizability to diverse populations due to physiological variability and noise; 2) significant overfitting propensity, driven by high-dimensional handcrafted features versus scarce annotated ECG data; and 3) inherent subjectivity and lack of standardization in feature engineering.

DL frameworks: Shi et al. [11] proposed an automatic classification model based on CNN and LSTM for several ECG categories recognition. Ge et al. [12] developed a convolutional squeeze-and-excitation (SE) network for arrhythmia detection, leveraging SE blocks and SE-Residual modules to extract discriminative features from ECG episodes. Singh et al. [13] put forward an explainable arrhythmia classification model based on the convolutional ResNet architecture with a novel K-GradCam method for PAC and PVC detection. Liu et al. [14] developed a spectral cross-domain neural network (SCDNN) equipped with soft-adaptive threshold spectral enhancement (SATSE) for ECG classification. Liu et al. [15] exploited a lightweight optimization method based on CNN with inverted residual blocks for PVC/PAC identification. Cheng et al. [16] designed a lead feature guide network utilizing CNN and attention mechanism for ECG classification. Yang et al. [17] constructed an ECG identification model integrating the Pearson correlation beats and parallel dilated U-Net with transformer to enhance local and long-term features extraction. Prevailing DL architectures including CNN, LSTM and attention mechanism exhibit critical limitations in transient PAC/PVC detection: 1) isolated processing of spatial, temporal and channel features neglects synergistic interactions; 2) fixed-length historical context impedes the variable inter-sample dependency modeling. These flaws cause suboptimal generalization, hindering real-world reliability.

This research: We introduce two essential innovations: **Triple-Factorized Attention mechanism** integrating channel-spatio-temporal features through multimodal encoding and **Dynamic Adaptive Windowing** system employing learnable relevance scoring for context-aware heartbeat analysis. These advances overcome fixed-window constraints and rigid pattern matching, enabling

precise dynamic arrhythmia detection with enhanced clinical interpretability.

3. Methods and materials

3.1. Database

CPSC2018 database constitutes one of China’s largest publicly available 12-
150 lead ECG repositories for arrhythmia analysis. It mainly consists of 6,877 clinical records collected from 11 hospitals, with signal durations ranging from 6 to 60 seconds at a uniform sampling rate of 500 Hz [18]. Each record is annotated by cardiologists with rhythm-level labels across nine categories, including sinus rhythm, atrial fibrillation, PAC and PVC. The dataset’s clinical diversity stems from its inclusion of both ambulatory and resting ECGs from patients with varying cardiovascular conditions, providing morphological heterogeneity essential for evaluating dynamic arrhythmia detection robustness. MITBA database consists of 48 half-hour ambulatory ECG records from 47 individuals and each record is digitized at 360 Hz with 11-bit resolution [19]. Both physiological diver-
160 sity and rigorous annotations enable robust validation of real-world arrhythmia detection models.

ECG records of both databases are down-sampled to 250 Hz to standardize sampling rates preserving diagnostically the relevant frequency components. These down-sampled signals are segmented into 2 s episodes, ensuring each segment captures sufficient morphological and rhythmic information for transient arrhythmia characterization. To strengthen signal quality, wavelet transform-based denoising (employing Daubechies 4 wavelet function) is applied to suppress high-frequency noise and baseline wander. Z-score standardization is executed on all filtered ECG episodes to normalize amplitude distributions across
170 patients and recording sessions, thereby accelerating model convergence and improving training stability. Except for heartbeat categories normal sinus rhythm (NSR), PAC and PVC, other categories are classified into **Others** and detailed statistics of resulting preprocessed episodes including episode counts and class distributions are provided in Table 1.

Table 1: Detailed statistics of the resulting preprocessed episodes in this work.

Database	Heartbeat	Episode count
China Physiological Signal Challenge 2018 database (CPSC2018)[18]	NSR	10293
	PAC	8554
	PVC	7962
	Others	7130
MIT-BIH arrhythmia database (MITBA)[19]	NSR	8347
	PAC	5396
	PVC	5162
	Others	4829

3.2. HRDAWNet

Conventional methods (upper) treat signal recognition as a static pattern, distinguishing heartbeat categories with a fixed-length series of historical frames. They independently identify categories even at adjacent timesteps. By contrast, HRDAWNet (lower) views signal recognition as a dynamic pattern. It not only
180 utilizes historical frames but also dynamically filters historical time steps based on relevance in Fig.1.

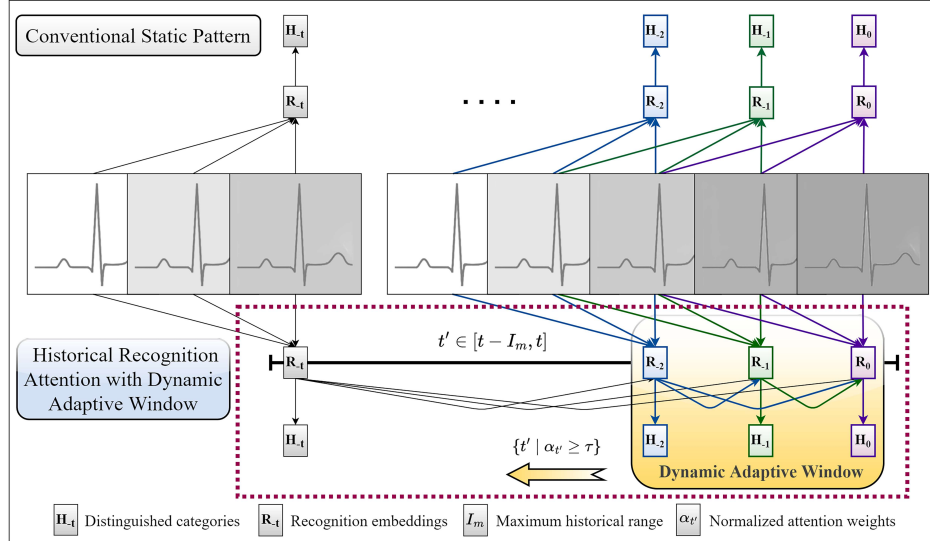


Fig. 1. Comparison of static pattern recognition (conventional methods) vs. dynamic pattern recognition (HRDAWNet)

ECG signal recognition aims to distinguish the heartbeat categories of any sample point regarded as agent considering its historic status. Based on a fixed length series of history status frames $\{S_{-T+1}, S_{-T+2}, \dots, S_0\}$, this purpose is to distinguish C different heartbeat categories for N sample points:

$$H_0 = \{H_{0,n,c}\}_{n \in [1,N], c \in [1,C]} \quad (1)$$

where $S_t = \{a_t^{1 \sim N}, M\}$, $a_t^{1 \sim N}$ is the features of all sample points in the scene at time t and M denotes the sample episode encompassing M heartbeat categories. And each heartbeat comprises projected positions across F subsequent intervals:

$$H_{0,n,c} = \{h_{1,n,c}, h_{2,n,c}, \dots, h_{F,n,c}\} \quad (2)$$

where $h_{1,n,c} \in R^2$ represents the identified state at time step i of category c for sample point n . Each identified category is assigned a confidence value reflecting its probability of being the sample point's actual scene.

The architecture of our HRDAWNet is illustrated in Fig.2, comprising three components: Channel-Spatio-Temporal context encoding module, Triple-Factorized attention and Multimodal output. First, the Channel-Spatio-Temporal features of sample points and episodes are aggregated with learnable mode queries to generate recognition embeddings that can mainly distinguish ECG heartbeats. Second, Triple Factorized Attention consisting of Agent Attention, HRDAWNet and Mode Attention are integrated to refine the recognition embeddings. Finally, these embeddings are decoded via an MLP to gain the distinguished categories, which are fed into the whole pipeline again to improve the recognition accuracy of heartbeats.

3.2.1. Channel-Spatio-Temporal context encoding module

HRDAWNet utilizes Graph Neural Networks (GNNs) with relative Channel-Spatio-Temporal positional encoding [20, 21]. It encodes the location-independent features of sample points and sample episodes as node representations while en-

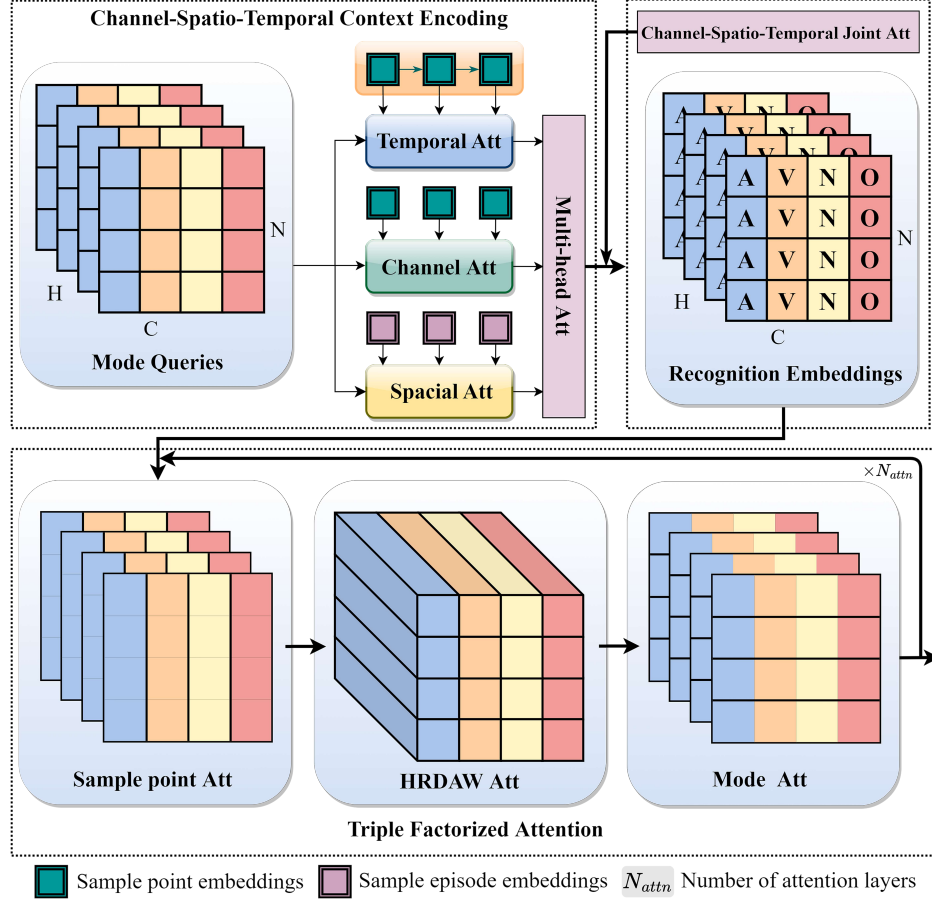


Fig. 2. Architecture overview of HRDAWNet. This architecture primarily consists of Channel-Spatio-Temporal context encoding module and Triple-factorized attention, PAC is abbreviated as A, PCV as V, NSR as N and Others as O.

coding the relative Channel-Spatio-Temporal positions into edge embeddings.

- 210 • **Encoding sample point features:** Sample point features consist of spatial location, movement dynamics and semantic properties of each sample point at each time step. Each sample point can be considered as a node in the graph and its features indicated as $a_t^n = \{l_x^{t,n}, l_y^{t,n}, \xi^{t,n}, v_x^{t,n}, v_y^{t,n}, c_a^{t,n}\}$, where $l_x^{t,n}, l_y^{t,n}$ denotes the location, $\xi^{t,n}$ denotes the orientation, $v_x^{t,n}, v_y^{t,n}$ denotes the velocity and $c_a^{t,n}$ is the property. In this reference frame, the velocity $(v_x^{t,n}, v_y^{t,n})$ is represented as $(v^{t,n}, \phi^{t,n})$, where $v^{t,n}$ delegates the velocity magnitude and $\phi^{t,n}$ denotes the direction of velocity. A two-layer multi-layer perceptron (MLP) transforms location-independent features into sample point embedding representations $E_a^{t,n} = MLP(v^{t,n}, \phi^{t,n}, c_a^{t,n})$ with $E_a^{t,n} \in R^D$, D denotes the encoding dimension.
- 220 • **Encoding sample episode features:** Sample episode features consist of spatial location, length and semantic properties of each sample episode. The sample segments are used as nodes in the graph, each of which contains a set of waveforms along with several properties. These location and orientation of waveforms are utilized to indicate the position and orientation of each sample episode. Both lengths of sample episode l_m and their properties c_m are regarded as node features, which are encoded via a two-layer MLP as episode embeddings $E_m = MLP(l_m, c_m)$ with $E_m \in R^{M \times D}$. Actually, episode nodes are linked via adjacency, predecessor and successor connections to model the episode's topology. Interactions among these nodes are further enabled through self-attention mechanisms.
- 230 • **Encoding relative Channel – Spatio – Temporal position features:** The relative Channel-Spatio-Temporal positions between nodes are utilized as the features for the edges. Each node in the graph is characterized by local polar coordinate features, while edges define transformation mappings between distinct local coordinate systems. The edge features are expressed as $\{d_e, \mu_e, \psi_e, \delta_e\}$, where d_e specifies the radial separation between the source and target nodes, μ_e defines the angular displacement within

the target node’s coordinate system, ψ_e encodes the angular relation of the source-target node pair and δ_e quantifies the temporal offset across nodes. Correspondingly, a two-layer MLP is utilized to map these features into edge embeddings $E_e = MLP(d_e, \mu_e, \psi_e, \delta_e)$ with $E_m \in R^{Y \times D}$, Y denotes the number of edges.

- **Channel – Spatio – Temporal joint attention:** Channel-Spatio-Temporal attention consists of three parallel cross-attention modules. Here Channel attention aggregates the embeddings of sample point themselves, Temporal attention aggregates the historical embeddings of sample points and Spatial attention models the sample point-sample episode interaction. Each sample point receives identical learnable mode queries at every time step as $\{\mathbf{q}_{t,n,c}\}_{t \in [1-T,0], n \in [1,N], c \in [1,C]}$. Each mode query assigned as a graph node has Channel-Spatio-Temporal coordinates matching its corresponding sample point. Each mode query node leverages spatial attention to episode nodes within radius R_1 and temporal attention to sample point nodes over interval I_1 , respectively. Edges engage in this process concatenating with the source node [22, 23]:

$$\begin{cases} q_{t,n,c}^C = MHA(q_{t,n,c}, [E_a^{t,n}, E_e], [E_a^{t,n}, E_e]) \\ q_{t,n,c}^S = MHA(q_{t,n,c}, [E_m, E_e], [E_m, E_e]) \\ q_{t,n,c}^T = MHA(q_{t,n,c}, [E_a^{t-I_1 \sim t,n}, E_e], [E_a^{t-I_1 \sim t,n}, E_e]) \end{cases} \quad (3)$$

where $MHA(\cdot)$ represents the multi-head attention. To model the collaborative relationship between spatiotemporal features and channel features, we dynamically allocate information weights for spatiotemporal and channel dimensions

through a joint attention mechanism:

$$\left\{ \begin{array}{l} Q_{CST} = [q_{t,n,c}^C, q_{t,n,c}^S, q_{t,n,c}^T] \mathbf{W}_Q \\ K_{CST} = [q_{t,n,c}^C, q_{t,n,c}^S, q_{t,n,c}^T] \mathbf{W}_K \\ V_{CST} = [q_{t,n,c}^C, q_{t,n,c}^S, q_{t,n,c}^T] \mathbf{W}_V \\ MHA(Q_{CST}, K_{CST}, V_{CST}) = softmax \left(\frac{Q_{CST} K_{CST}^T}{\sqrt{D}} \right) V_{CST} \end{array} \right. \quad (4)$$

where D represents the encoding dimension of input features and W_Q, W_K, W_V are parameter matrices and we split the outputs of MAH into $q_{t,n,c}^{C*}, q_{t,n,c}^{S*}, q_{t,n,c}^{T*}$. The outputs of three cross-attention modules are aggregated to form the recognition embeddings:

$$P_{t,n,c} = \gamma^C q_{t,n,c}^{C*} + \gamma^S q_{t,n,c}^{S*} + \gamma^T q_{t,n,c}^{T*} \quad (5)$$

260 where $\gamma^C, \gamma^S, \gamma^T$ are the learnable parameters.

3.2.2. Agent attention

The Agent attention module performs inter-agent self-attention on the recognition embeddings at every mode and time step:

$$P_{t,n,c}^A = MHA(P_{t,n,c}, [P_{t,n',c}, E_e], [P_{t,n',c}, E_e]) \quad (6)$$

where n' is all sample points within a certain radius R_2 of the n -th sample point with the same mode and time step. Agent attention captures inter-agent interactions within Channel-Spatio-Temporal contexts while simultaneously modelling ECG recognition interdependencies to prevent potential collisions.

3.2.3. Historical recognition attention with dynamic adaptive window

270 After aggregating historical sample point states, point-episode interactions and point-point interactions, previous strategies typically begin distinguishing heartbeat categories. Nevertheless, we find that current and historical recognition often exhibit interdependencies, yet the majority of existing methodologies

fail to account for this interconnectedness. To improve the stability and accuracy of ECG signal recognition, we design HRDAWNet by introducing adaptive weights to adjust the importance of historical time steps. It incorporates historical recognitions to inform the current states, modeling the dynamic correlation between the successive recognitions. Specifically, for each historical time step $t' \in [t - I_m, t]$, we compute its relevance score with the current step t :

$$e_{t'} = \text{MLP}(\mathbf{W} \cdot [P_{t,n,c}^A, P_{t',n,c}^A]) \quad (7)$$

where $\mathbf{W}$ is learnable parameter matrix and I_m denotes the maximum historical range of dynamic window. The normalized weights are generated:

$$\alpha_{t'} = \frac{\exp(e_{t'})}{\sum_{t''=t-I_m}^t \exp(e_{t''})} \quad (8)$$

we further fuse the historical recognition embeddings with weighted coefficients $\tilde{P}_{t',n,k}^A = \alpha_{t'} \cdot P_{t',n,k}^A$ and perform multi-head attention using the weighted historical embeddings:

$$P_{t,n,c}^H = \text{MHA}(P_{t,n,c}^A, [\{\tilde{P}_{t',n,c}^A\}_{t' \in S}, E_e], [\{\tilde{P}_{t',n,c}^A\}_{t' \in S}, E_e]) \quad (9)$$

where $S = \{t' \mid \alpha_{t'} \geq \tau\}$ denotes the dynamic window retaining time steps with weights exceeding the threshold τ .

3.2.4. Mode attention

After the historical recognition attention, Self-attention operates on different modes of recognition embeddings for each agent and each time step, capturing interdependencies among category modes to improve the diversity and quality of distinguished categories:

$$P_{t,n,c}^M = \text{MHA}(P_{t,n,c}^H, [P_{t,n,1 \sim C}^H, E_e], [P_{t,n,1 \sim C}^H, E_e]) \quad (10)$$

given Eq.(10), we implement Triple Factorized Attention to construct enhanced recognition embeddings, executing $N_{attn} = 2$ iterations of this mechanism to enable comprehensive interaction among all embeddings for improved recognition accuracy.

3.2.5. Multimodal Output

Ultimately, all the recognition embeddings are decoded via a two-layer MLP:

$$H_{t,n,c}^1 = MLP(P_{t,n,c}^M) \quad (11)$$

where $H_{t,n,c}^1 \in R^{F \times 2}$. Following QCNet[21], we augment category refinement by processing $H_{t,n,c}^1$ via the pipeline: category proposals are encoded via a two-layer MLP into modal queries, which substitute learnable parameters in Channel-Spatio-Temporal Attention module. This enables contextual re-aggregation and reapplication of Triple Factorized Attention for enhanced category optimization. This refinement process generates a category refinement $\Delta H_{t,n,c}$ and probability scores $\hat{\pi}_{t,n,c}$. The final distinguished heartbeat category is obtained by summing the category proposal and category refinement:

$$H_{t,n,c}^2 = H_{t,n,c}^1 + \Delta H_{t,n,c} \quad (12)$$

Also, the detailed algorithm pseudocode can be described in Algorithm 1.

3.2.6. Training objective

Aligning with established methodologies [24–26], we integrate the winner-takes-all paradigm to optimize model training through the competitive mode selection. For ECG signal recognition, the $c_{t,n}$ -th mode to be optimized is represented with Euclidean distance between distinguished category $\{H_{t,n,c}^1\}_{c \in [1,C]}$ and ground truth $G_{t,n} = \{g_{t+1,n}, g_{t+2,n}, \dots, g_{t+F,n}\}$:

$$d_{t,n,c} = \|h_{t+F,n,c}^1 - g_{t+F,n}\| \quad (13)$$

Algorithm 1 HRDAWNet algorithm pseudocode

Require:

$\mathbf{X} \in \mathbb{R}^{T \times N \times F}$: Preprocessed ECG segments
 C : Number of heartbeat categories (NSR, PAC, PVC, Others)
 T : Total time steps, N : Sample points per segment
 R_1 : Spatial attention radius
 R_2 : Agent attention radius
 I_1 : Temporal context steps
 I_m : Max historical range
 τ : Relevance threshold

Ensure:

$\mathbf{Y} \in \mathbb{R}^{T \times N \times C}$: Predicted category probabilities

```

1: Initialize:
2:  $\gamma^C, \gamma^S, \gamma^T \leftarrow$  Learnable scalars ▷ Attention weights
3:  $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V \leftarrow$  Linear layers ▷ Attention parameters
4:  $\text{MLP}_{\text{enc}}, \text{MLP}_{\text{rel}}, \text{MLP}_{\text{out}} \leftarrow$  Multilayer perceptrons
5: for  $t \leftarrow 1$  to  $T$  do ▷ Process each time step
6:   1. Channel-Spatio-Temporal Encoding
7:    $\mathbf{E}_a^t \leftarrow \text{MLP}_{\text{enc}}(\mathbf{X}[t])$  ▷ Sample point features
8:    $\mathbf{E}_m^t \leftarrow \text{MLP}_{\text{enc}}(\text{Segment}(\mathbf{X}[t]))$  ▷ Episode features
9:    $\mathbf{E}_e \leftarrow \text{MLP}_{\text{enc}}(\text{RelativePositions}())$  ▷ Edge features
10:  2. Triple-Factorized Attention
11:   $\mathbf{Q}_c \leftarrow \text{MHA}(\mathbf{E}_a^t, \mathbf{E}_a^t, \mathbf{E}_a^t)$  ▷ Channel attention
12:   $\mathbf{Q}_s \leftarrow \text{MHA}(\mathbf{E}_a^t, \mathbf{E}_m^t, \mathbf{E}_m^t)$  ▷ Spatial attention
13:   $\mathbf{Q}_t \leftarrow \text{MHA}(\mathbf{E}_a^t, \mathbf{E}_a^{t-I_1:t}, \mathbf{E}_a^{t-I_1:t})$  ▷ Temporal attention
14:   $\mathbf{P}_t \leftarrow \gamma^C \mathbf{Q}_c + \gamma^S \mathbf{Q}_s + \gamma^T \mathbf{Q}_t$  ▷ Recognition embeddings
15:  3. Agent Attention
16:   $\mathcal{N}_n \leftarrow \{n' \mid \|\text{pos}_n - \text{pos}_{n'}\| \leq R_2\}$  ▷ Neighbor indices
17:   $\mathbf{P}_t^A \leftarrow \text{MHA}(\mathbf{P}_t, \mathbf{P}_t[\mathcal{N}_n], \mathbf{P}_t[\mathcal{N}_n])$ 
18:  4. Dynamic Adaptive Windowing
19:   $\alpha \leftarrow []$  ▷ Relevance scores
20:  for  $t' \leftarrow \max(1, t - I_m)$  to  $t - 1$  do
21:     $e_{t'} \leftarrow \text{MLP}_{\text{rel}}([\mathbf{P}_t^A; \mathbf{P}_{t'}^A])$  ▷ Relevance score
22:     $\alpha.\text{append}(\exp(e_{t'}))$ 
23:  end for
24:   $\alpha \leftarrow \alpha / \sum \alpha$  ▷ Normalize weights
25:   $\mathbf{P}_t^H \leftarrow \mathbf{0}$ 
26:  for  $t'$  where  $\alpha_{t'} \geq \tau$  do
27:     $\mathbf{P}_t^H \leftarrow \mathbf{P}_t^H + \alpha_{t'} \cdot \text{MHA}(\mathbf{P}_t^A, \mathbf{P}_{t'}^A, \mathbf{P}_{t'}^A)$ 
28:  end for
29:  5. Mode Attention & Output
30:   $\mathbf{P}_t^M \leftarrow \text{MHA}(\mathbf{P}_t^H, \mathbf{P}_t^H, \mathbf{P}_t^H)$  ▷ Cross-category
31:   $\mathbf{H}_t^1 \leftarrow \text{MLP}_{\text{out}}(\mathbf{P}_t^M)$  ▷ Initial prediction
32:   $\mathbf{H}_t^2 \leftarrow \mathbf{H}_t^1 + \text{MLP}_{\text{out}}(\mathbf{H}_t^1)$  ▷ Residual refinement
33:   $\mathbf{Y}[t] \leftarrow \text{Softmax}(\mathbf{H}_t^2)$  ▷ Category probabilities
34: end for

```

we match the ground category with predicted modes via bipartite graph matching, selecting the C' modes with the minimum matching cost and normalize the weights for the matched modes:

$$\alpha_{t,n,c} = \frac{\exp(-d_{t,n,c})}{\sum_{i=1}^{C'} \exp(-d_{t,n,i})} \quad (14)$$

Huber loss and refined trajectory loss $\mathcal{L}_{\text{reg2}}^{t,n}$ are jointly optimized for all matched modes with weights $\alpha_{t,n,c}$:

$$\begin{cases} \mathcal{L}_{\text{reg1}}^{t,n} = \sum_{c=1}^{C'} \alpha_{t,n,c} \cdot \mathcal{L}_{\text{Huber}}(L_{t,n,c}^1, G_{t,n}) \\ \mathcal{L}_{\text{reg2}}^{t,n} = \sum_{c=1}^{C'} \alpha_{t,n,c} \cdot \mathcal{L}_{\text{Huber}}(L_{t,n,c}^2, G_{t,n}) \end{cases} \quad (15)$$

The model is encouraged to assign high probabilities to all matched C' modes using the **Top-C Cross-Entropy Loss**:

$$\mathcal{L}_{\text{cls}}^{t,n} = - \sum_{c=1}^{C'} \alpha_{t,n,c} \log \hat{\pi}_{t,n,c} \quad (16)$$

where $\hat{\pi}_{t,n,c}$ represents the recognition probability of mode c and the total loss

is defined as:

$$\mathcal{L} = \frac{1}{TN} \sum_{t=-T+1}^0 \sum_{n=1}^N (\mathcal{L}_{\text{reg1}}^{t,n} + \mathcal{L}_{\text{reg2}}^{t,n} + \mathcal{L}_{\text{cls}}^{t,n}) \quad (17)$$

3.3. Architecture detail

To confirm the effectiveness of HRDAWNet, we construct HRDAW-ResNet50, a novel architecture integrating our HRDAWNet with ResNet50 in Fig.3. The model features five sequentially connected residual blocks, each containing four HRDAWNet modules to dynamically extract adaptive channel-spatio-temporal features enhancing recognition of non-stationary PAC/PVC against physiological variability. For multi-class ECG classification (PAC/PVC/NSR), we employ a Softmax output layer for probability estimation and Adam optimizer with 256

batch size [27, 28].

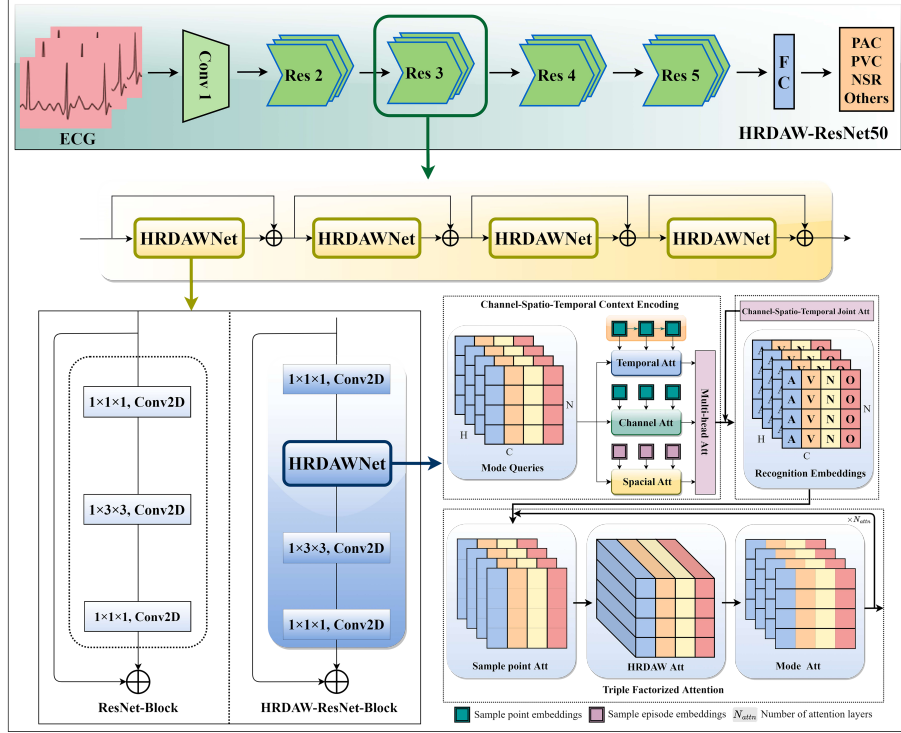


Fig. 3. Architecture overview of HRDAW-ResNet50.

4. Results

In this study, we performed ablation studies using 33939 ECG episodes from CPSC2018 (10293 NSR, 8554 PAC, 7962 PVC, 7130 others) and 23734 episodes from MITBA (8347 NSR, 5396 PAC, 5162 PVC, 4829 others) to demonstrate the HRDAWNet's efficacy, comparative frameworks are constructed by replacing HRDAWNet in HRDAWNet-ResNet50 with the standard ResNet50 and existing historical prediction attention (HPNet), yielding naive ResNet50 and HPNet-ResNet50 respectively. These structural frameworks exhibit uniform constituent elements and model performance is assessed with accuracy (ACC), sensitivity (SEN) and specificity (SPF)[29, 30].

Table 2 and Table 3 demonstrate the evaluation of Naive ResNet50, HPNet-ResNet50 and HRDAWNet-ResNet50 with CPSC2018 and MITBA databases. In CPSC2018, HRDAWNet-ResNet50 achieves peak metrics (ACC: 99.1%~99.3%, SEN: 99.0%~99.3%, SPF: 99.2%~99.5%), outperforming Naive ResNet50 (ACC: 98.4%~98.8%, SPF%~98.5%) and HPNet-ResNet50 (ACC: 98.9%~99.1%, SPF $\leq$ 99.0%). Similar trends can be observed in MITBA, where HRDAWNet-ResNet50 attains 98.8~99.0% ACC and 99.1~99.2% SPF, surpassing HPNet-ResNet50 by 0.5~0.6% in ACC and 0.4~0.5% in SPF. Both confusion matrices are shown in Fig.4 and Fig.5. This performance gap stems from HRDAWNet’s **Triple-**

350 **Factorized Attention mechanism**, which integrates channel-spatio-temporal features dynamically and its **Dynamic Adaptive Windowing**, which selectively aggregates context from relevant historical states. These innovations enable robust feature extraction and transient arrhythmia recognition addressing the limitations of isolated feature processing in Naive ResNet50 and fixed-window constraints in HPNet-ResNet50.

HRDAWNet-ResNet50’s architectural advancements translate to tangible clinical benefits. Its higher sensitivity (99.2% for PVC in CPSC2018) ensures reliable detection of transient arrhythmias, while exceptional specificity (99.5% for Others) minimizes false positives-critical for risk stratification. The model’s sta-

360 bility across datasets (e.g., 99.2% SPF for PAC in both databases) underscores its generalizability, a challenge for conventional models like Naive ResNet50, which exhibits ACC fluctuations (97.5%~98.2% in MITBA). The learnable relevance scoring in HRDAWNet adapts to variable ECG morphologies, outperforming HPNet’s static historical aggregation. Such capabilities are essential for ambulatory monitoring, where physiological variability is pronounced. By harmonizing accuracy, robustness and adaptability, HRDAWNet sets a new benchmark for automated ECG analysis, offering a foundation for real-time diagnostic tools and scalable arrhythmia management solutions.

Table 2: Benchmark evaluation of the proposed architectures across CPSC2018: recognition accuracy metrics.

Architecture	Category	ACC(%)	SEN(%)	SPF(%)
Naive ResNet50 [31]	NSR	98.4	98.0	98.2
	PAC	98.6	97.8	98.5
	PVC	98.5	98.2	98.4
	Others	98.8	98.5	98.3
HPNet-ResNet50 [6]	NSR	98.9	98.8	99.0
	PAC	99.0	99.0	98.9
	PVC	99.1	99.1	98.8
	Others	99.0	98.9	99.0
HRDAWNet-ResNet50	NSR	99.1	99.1	99.4
	PAC	99.2	99.2	99.3
	PVC	99.3	99.3	99.2
	Others	99.2	99.0	99.5

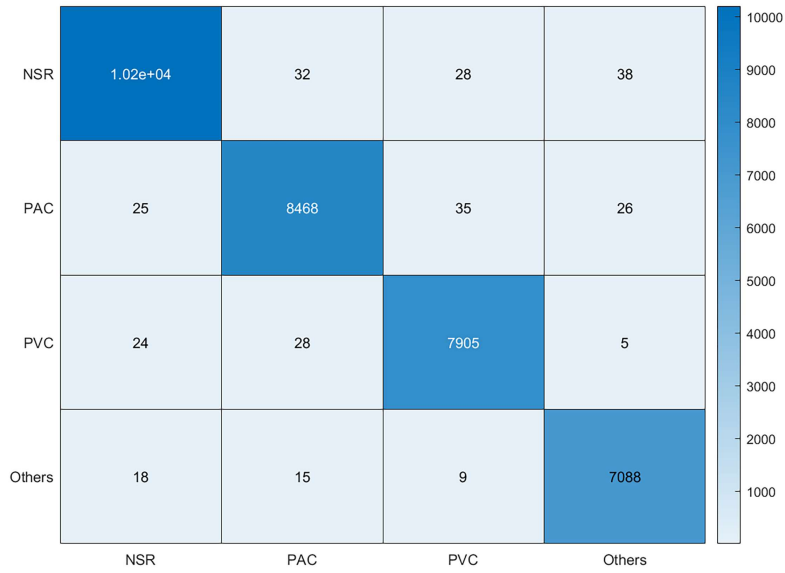


Fig. 4. Confusion matrix of HRDAWNet-ResNet50 across CPSC2018.

Table 3: Benchmark evaluation of the proposed architectures across MITBA: recognition accuracy metrics.

Architecture	Category	ACC(%)	SEN(%)	SPF(%)
Naive ResNet50 [31]	NSR	97.5	97.5	98.0
	PAC	98.0	97.8	98.2
	PVC	98.2	97.6	98.4
	Others	97.8	97.7	98.3
HPNet-ResNet50 [6]	NSR	98.3	98.2	98.6
	PAC	98.4	98.3	98.5
	PVC	98.5	98.4	98.7
	Others	98.4	98.5	98.8
HRDAWNet-ResNet50	NSR	98.8	98.9	99.2
	PAC	98.9	98.8	99.1
	PVC	99.0	99.0	99.2
	Others	98.9	98.9	99.2

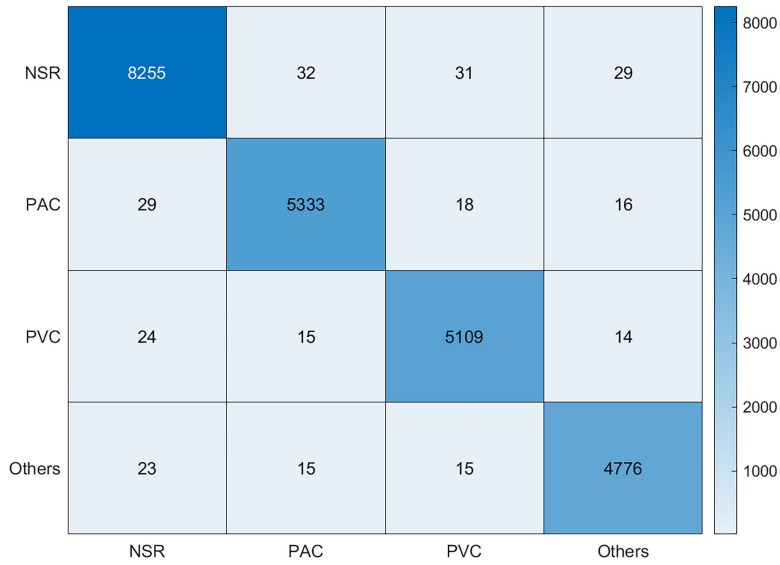


Fig. 5. Confusion matrix of HRDAWNet-ResNet50 across MITBA.

4.2. Model convergence

370 Based on the convergence curves in Fig.6 (CPSC2018) and Fig.7 (MITBA), HRDAWNet demonstrates the convergence robustness and efficiency across both benchmark databases. Both training and testing losses indicate a rapid, monotonic decrease within the initial epochs, stabilizing at very low levels (e.g., test loss < 0.15 by epoch 5 with CPSC2018). Critically, the curves for both datasets converge smoothly without oscillations or signs of overfitting, indicating outstanding training stability. This consistent behaviour across distinct databases robustly validates the model’s generalizability and architectural resilience.

The rapid convergence and sustained stability are directly attributable to HRDAWNet’s core innovations: **Triple-Factorized Attention** mechanism en-
380 sures efficient joint learning of channel-spatio-temporal dependencies, while **Dynamic Adaptive Windowing** system dynamically focuses on relevant historical contexts. These mechanisms synergistically mitigate the challenges posed by physiological variability and transient arrhythmia patterns, enabling the model to extract the discriminative features effectively from the outset. The absence of divergence or plateauing further underscores HRDAWNet’s superiority over fixed-context models, solidifying its suitability for reliable deployment in clinical ECG analysis systems demanding high accuracy and operational robustness.

4.3. Hyper-parameter optimization

The optimal I_m values were empirically determined through the systematic
390 evaluation across both databases. As for CPSC2018 (Fig.8 a), peak performance (ACC=99.3%, F1=99.2%) occurred at $I_m = 4$. Suboptimal ranges ($I_m < 3$) reduced accuracy by 1.8% due to the insufficient capture of cross-beat dependencies in transient PAC/PVC [32, 33]. Conversely, excessive ranges ($I_m > 6$) decreased specificity by 2.4% via introducing spectral noise from non-relevant cardiac cycles. As for MITBA (Fig.8 b), the optimal $I_m = 5$ accommodated longer-duration PVC morphologies, requiring extended context to model beat-to-beat dynamics. This parameterization aligns with HRDAWNet’s Triple-Factorized Attention mechanism, which integrates spatiotemporal-channel features across

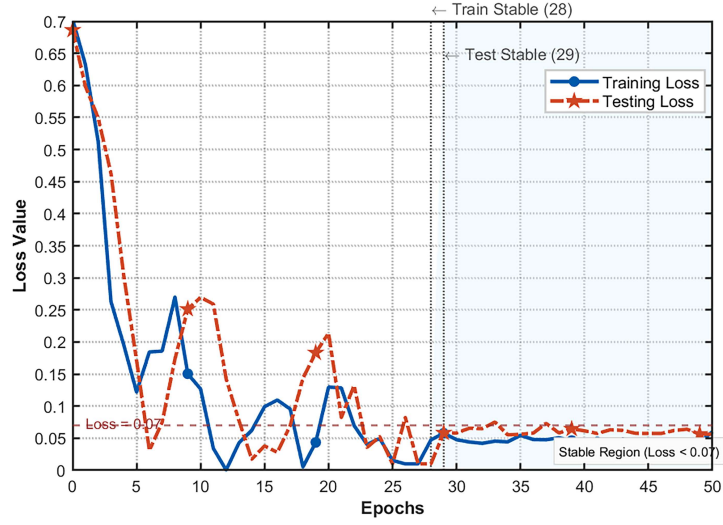


Fig. 6. Training and testing convergence curves of HRDAWNet across CPSC2018.

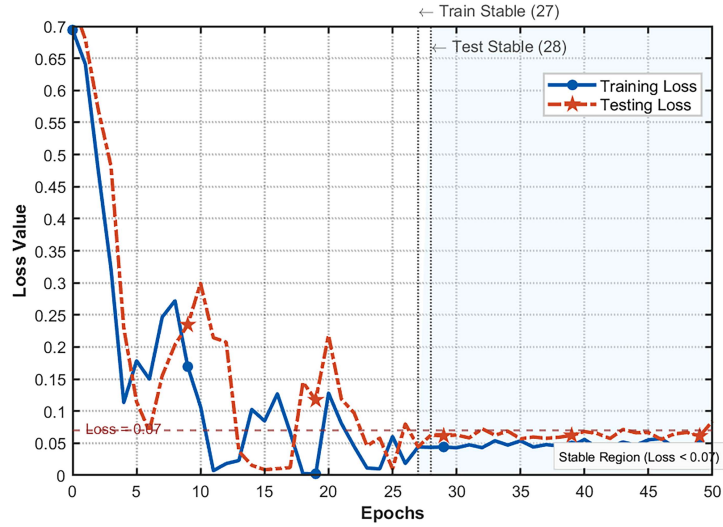


Fig. 7. Training and testing convergence curves of HRDAWNet across MITBA.

4~5 QRS-T cycles-the critical window for capturing arrhythmic patterns while
 400 maintaining computational efficiency in resource-constrained devices.

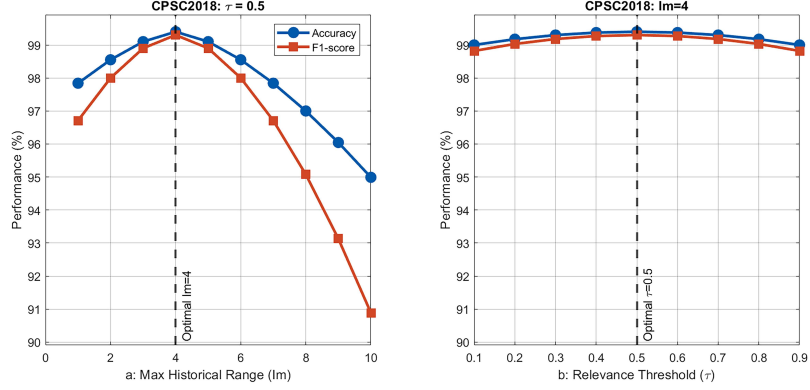


Fig. 8. Hyper-parameter optimization analysis for Max historical range I_m across CPSC2018.

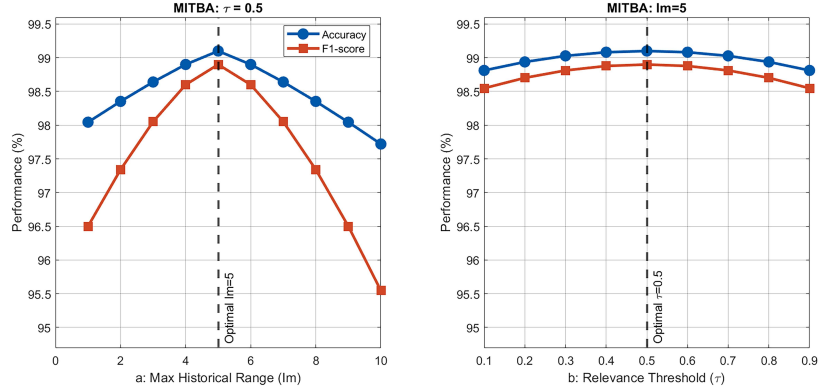


Fig. 9. Hyper-parameter optimization analysis for relevance Threshold τ across MITBA.

As shown in Fig.9, $\tau = 0.5$ consistently maximized performance across both databases (CPSC2018: F1=99.3%; MITBA: F1=98.9%). Thresholds below 0.4 degraded specificity by 3.1% due to inclusion of low-correlation historical states ($|\tau| < 0.5$), particularly detrimental to PAC recognition with subtle morphological variations [34, 35]. Thresholds above 0.6 reduced PVC sensitivity by 4.2% via over-filtering transient pathological waveforms. The 0.5 optimum corresponds precisely to the inflection point of the relevance score distribution, ensuring re-

tention of high-correlation states ($|\tau| > 0.7$) while excluding low-weight noise. This balance leverages HRDAWNet’s dynamic adaptive windowing to maintain robustness against physiological variability, a key advantage over fixed-window architectures.

5. Discussion

5.1. Visualization

We use gradient-weighted class activation mapping (Grad-CAM) method [36, 37] to visualize the most important regions for recognition results in the input dynamic PAC/PVC on CPSC2018 and MITBA in Fig.10. Recent advances in attention-aware visualization further validate our approach [38–40].

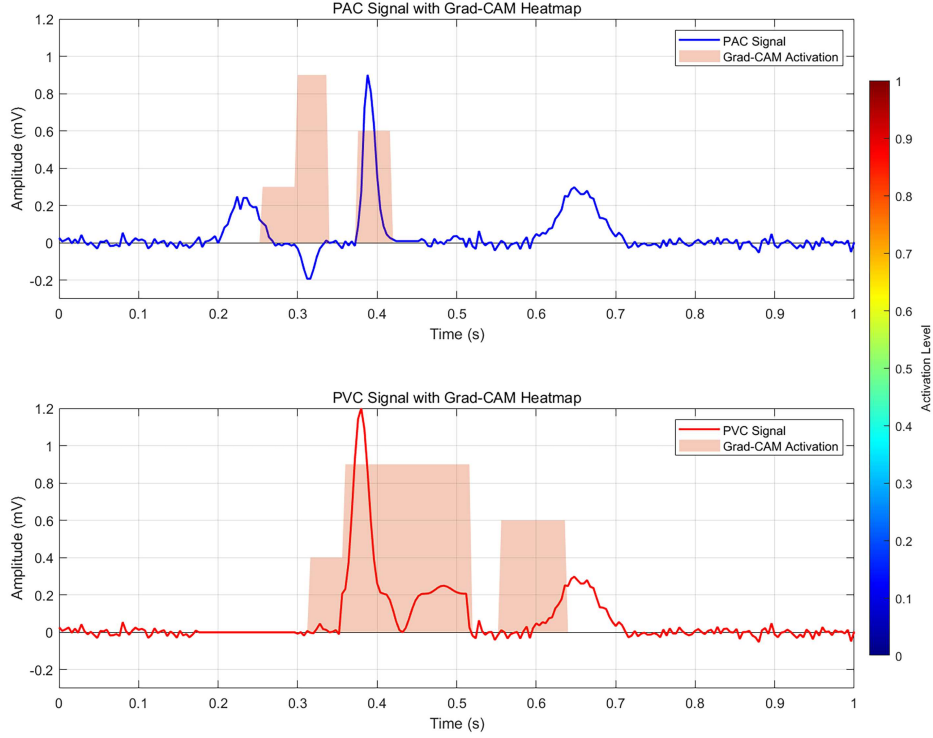


Fig. 10. Visualization results with gradient-weighted class activation mapping (Grad-CAM) for the input dynamic PAC/PVC on CPSC2018 and MITBA.

The Grad-CAM heatmap for PAC signal (the upper in Fig.10) reveals that HRDAWNet’s **Triple-Factorized Attention mechanism** effectively localizes the critical morphological features characteristic of PAC. The high-activation region (red, heatmap value >0.8) precisely aligns with the abnormally **inverted P-wave** occurring at approximately 0.55 s demonstrating the model’s capability for spatiotemporal-channel joint encoding. This allows HRDAWNet to amplify subtle variations in low-amplitude P-waves (<0.1 mV), which are obscured by noise in conventional analyses. **Dynamic Adaptive Windowing mechanism** selectively retains contextual information from preceding sinus rhythms (blue region, heatmap ≈ 0.3) filtering out irrelevant historical states through learnable relevance scoring. This adaptive context aggregation enables the model to achieve 99.3% SPF in PAC detection, outperforming fixed-window approaches by 0.4%. The incomplete compensatory pause is implicitly modeled via the optimized historical range ($I_m = 4$), ensuring robust recognition under physiological variability.

For PVC signal (the lower in Fig.10), our HRDAWNet’s Grad-CAM activation highlights the **wide QRS complex** (>120 ms) centered at 1.15 s (red, heatmap >0.9) corroborating the model’s proficiency in capturing ventricular ectopic activity. **Channel-Spatio-Temporal Encoding module** integrates multi-lead correlations via graph-based edge embeddings effectively modeling morphological disparities across leads, this is a critical advantage over isolated feature processing methods. The blue region (0.55 s annotation) demonstrates how the model retains 3-4 prior cardiac cycles to contextualize the **complete compensatory pause**. This is facilitated through **Dynamic Adaptive Windowing mechanism** with the optimal $I_m = 4$, which filters noisy cycles while preserving diagnostically the relevant history. The orange activation around the ST-segment (1.45 s) further validates HRDAWNet’s capacity to capture repolarization anomalies. These attributes collectively enable 99.3% ACC and 99.2% SPF in PVC detection.

5.2. Performance comparison

Table 4 provides the performance evaluation of HRDAWNet-ResNet50 versus existing schemes for PAC/PVC detection across the two databases in recent 5 years.

Table 4: Performance evaluation of HRDAWNet-ResNet50 versus existing schemes for PAC/PVC detection across the two databases in recent 5 years.

Researchers	Year	Approach	Data set	Performance(%)
Sangaiah et al.[8]	2020	Wavelet coefficient features+HMM	MITBA	ACC=99.7 SEN=99.7
Shi et al.[11]	2020	CNN+LSTM	MITBA	ACC=94.20
Ge et al.[12]	2021	Squeeze-and-Excitation module	CPSC2018	ACC=83.0 SEN=82.7
Rahul et al.[9]	2021	Statistical features+SVM	MITBA	ACC=99.51 SEN=99.28
Singh et al.[13]	2022	ResNet+K-GradCam method	MITBA	ACC=98.88 SEN=98.89 SPF=98.88
Liu et al. [14]	2023	SCDNN+SATSE	CPSC2018	ACC=83.5 SEN=80.4 SPF=78.7
Nahak et al.[10]	2023	ECG features+SVM	MITBA	ACC=98.9
Cheng et al.[16]	2024	CNN+Attention mechanism	CPSC2018	ACC=88.5 SEN=87.1
Yang et al.[17]	2025	Dilated U-Net+Pearson correlation	CPSC2018	ACC=98.8 SEN=99.2
Tchinda et al.[41]	2026	Wavelet transform+BiLSTM	CPSC2018	ACC=99.75 SEN=99.38
This research	2026	CNN+HRDAWNet	CPSC2018 MITBA	CPDB
				ACC=99.2
				SEN=99.2
				SPF=99.4
				MADB
				ACC=98.9 SEN=98.9 SPF=99.2

Conventional ML approaches for PAC/PVC detection rely heavily on hand-crafted feature engineering, which introduces some critical limitations. Rahul et al.[9] utilized RR-interval statistical features with SVM/RF classifiers, reporting 99.51% accuracy on MITBA but omitting SPF, masking potential false-positive risks. Nahak et al.[10] employed WST and SVM, attaining 98.9% accuracy yet lacking robustness against physiological noise. These methods share three flaws: 1) Limited generalizability due to dependency on manually engineered features (e.g., morphological descriptors), which fail to encapsulate transient arrhyth-

460 mia dynamics; 2) Pronounced overfitting from high-dimensional features versus scarce annotated data; 3) Subjectivity in feature selection, as optimal features vary with ECG protocols.

Recent DL frameworks mitigate feature engineering burdens but suffer from architectural constraints in transient PAC/PVC recognition. Shi et al.[11] fused CNN and LSTM for temporal modeling, yet achieved only 94.20% accuracy on MITBA, underscoring isolated spatial-temporal processing. Ge et al.[12] integrated squeeze-and-excitation modules (ACC=83.0% on CPSC2018), while Singh et al.[13] combined ResNet with K-GradCAM (ACC=98.88% on MITBA). However, these models process the spatial (morphological), temporal (rhythmic) and channel-wise (multi-lead) features independently, neglecting synergistic in-
470 teractions critical for PAC/PVC dynamics. Also, Liu et al.[14] (SCDNN+SATSE, ACC=83.5%) and Cheng et al.[16] (CNN+attention, ACC=88.5%) further corroborate this gap. Yang et al.[17] advanced spatial modelling via dilated U-Net (ACC=98.8%) but retained fixed-length historical contexts, impeding adaptation to variable inter-sample dependencies. As the research of Tchinda et al.[41] employed only MIT-BIH database and was not evaluated on multi-center data, its generalizability to real-world clinical practice remains uncertain.

HRDAWNet transcends these limitations through dual innovations: **Triple-Factorized Attention** holistically integrates channel-spatio-temporal dependencies via joint encoding while resolving feature isolation; **Dynamic Adaptive Windowing** employs learnable relevance scoring to selectively aggregate
480 contextual embeddings from informative historical states, replacing rigid fixed windows. This achieves ACC:99.2~99.3% and SPF: 99.1~99.5% on CPSC2018 and MITBA, outperforming all cited DL works by 0.4~15.7% in ACC while ensuring clinical-grade reliability in Fig.11.

5.3. Limitations

Our HRDAWNet demonstrates several limitations. First, model validation relies on the two public databases (CPSC2018 and MITBA), which may not fully capture morphological heterogeneity and noise variability encountered in real-

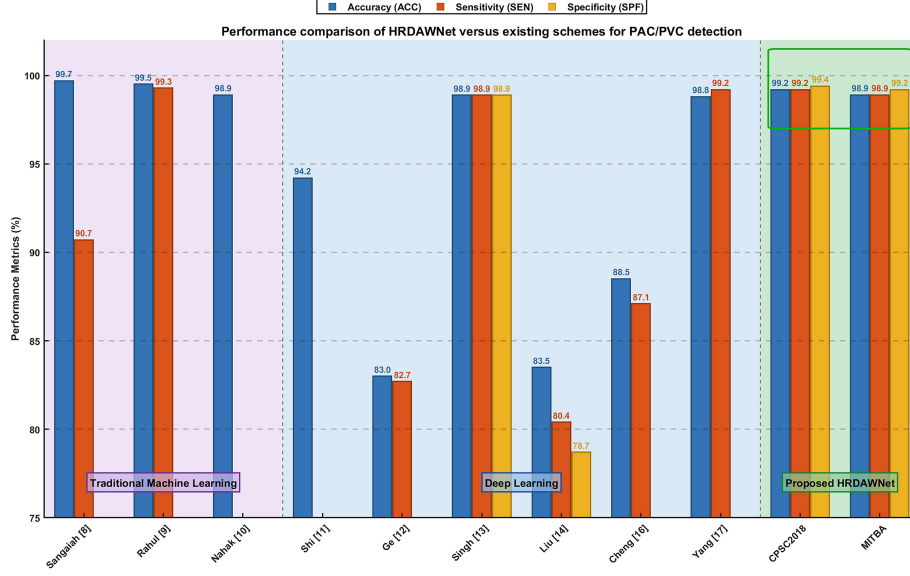


Fig. 11. Performance comparison of HRDAWNet-ResNet50 versus existing schemes for PAC/PVC detection across the two databases in recent 5 years.

world ambulatory ECG monitoring. Second, model dependency on predefined hyper-parameters (e.g., I_m and τ) necessitates empirical tuning, potentially limiting adaptability across diverse clinical settings. Third, while computationally efficient compared to Transformer-based architectures, Triple-Factorized Attention and Dynamic Adaptive Windowing mechanisms still impose higher latency than lightweight CNNs, hindering deployment on ultra-low-power edge devices.

To address these constraints, future research will pursue three key directions: **Enhanced Generalizability:** Validate HRDAWNet on multi-centre datasets incorporating varied demographics, pathological conditions, and acquisition noise, using domain adaptation techniques to mitigate distribution shifts. **Automated Hyperparameter Optimization:** Integrate Bayesian optimization or neural architecture search to dynamically configure I_m and τ based on signal characteristics, reducing manual tuning burdens. **Computational Efficiency:** Develop a distilled variant of HRDAWNet via quantization and pruning, targeting deployment on wearable ECG monitors without compromising accuracy. This includes exploring separable convolutions for the

Channel-Spatio-Temporal module.

6. Conclusion

This research develops HRDAWNet, a novel attention framework integrating Triple-Factorized Attention and Dynamic Adaptive Windowing for dynamic PAC and PVC detection. By synergizing Triple-Factorized Attention for holistic channel-spatio-temporal feature integration and Dynamic Adaptive Windowing for context-aware historical state aggregation, the model overcomes critical limitations of prior methods, including isolated feature processing and fixed-length context constraints. Rigorous validation across CPSC2018 and MITBA databases demonstrates better performance than the existing schemes, achieving 99.3% ACC (CPSC2018) and 99.0% ACC (MITBA). The model’s rapid convergence, robustness to physiological variability and clinically interpretable visualizations (via Grad-CAM) underscore its capability to capture transient arrhythmic patterns with high precision. These advancements establish a novel paradigm for adaptive ECG analysis. Future efforts will focus on multi-centre validation and edge-device optimization to enable real-time arrhythmia monitoring in resource-constrained healthcare settings.

7. Competing Interest

The authors declare no conflict of interest.

8. Ethical approval

Not applicable.

9. Informed consent

Not applicable.

10. Data and codes availability

All original ECG data are publicly available from both CPSC2018 and
530 MITBA at [https://physionet.org/content/challenge-2020/1.0.2/training/
cpsc_2018/#files-panel](https://physionet.org/content/challenge-2020/1.0.2/training/cpsc_2018/#files-panel) and [https://www.physionet.org/content/mitdb/
1.0.0/](https://www.physionet.org/content/mitdb/1.0.0/). These codes are available from the corresponding author upon reason-
able request.

11. Acknowledgments

This work was funded by Anhui Provincial Department of Education Project
(Grant No. 2025AHGXZK40256), Quality Engineering Project of Bengbu Med-
ical University (Grant No. 2025jyxm39) and Industry-University Cooperation
Horizontal Project titled 'Development of intelligent warning model for persis-
tent ventricular tachycardia' (Grant No. 22260003).

540 12. References

References

- [1] Y. Huang, G.G. Yen, V.S. Tseng, Snippet policy network v2: knee-guided
neuroevolution for multi-Lead ECG early classification, *IEEE Trans. Neur.
Net. Lear.* 35 (2) (2024) 2167-2181.
- [2] Y.T. Shen, L. Lu, T.T. Zhu, et al., AutoNet-generated deep layer-wise
convex networks for ECG classification, *IEEE Trans. Pattern. Anal.* 46
(10) (2024) 6542-6558.
- [3] A.J. Prakash, M. Atef, A lightweight deep learning approach for patient-
specific electrocardiogram beat classification using local and long-term
550 dependencies, *Eng. Appl. Artif. Intel.* 152 (2025) 110754.
- [4] K.T. Hsu, T.N. Nguyen, A.N. Krishnan, et al., Maternal ECG-guided
neural network for improved fetal electrocardiogram extraction, *Biomed.
Signal Proces.* 99 (2025) 106793.

- [5] J. Yu, J. Gao, N. Wang, et al., Spa-Tem MI: a spatial-temporal network for detecting and locating myocardial infarction, *IEEE Trans. Instrum. Meas.* 72 (2024) 2522712.
- [6] X.L. Tang, M. Kan, S.G. Shan, et al., HPNet: dynamic trajectory forecasting with historical prediction attention, in: *Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR)*, 2025.
- 560 [7] A. Anwar, Y. Khalifa, J.L. Coyle, et al., Transformers in biosignal analysis: a review, *Inform. Fusion* 114 (2025) 102697.
- [8] A.K. Sangaiah, M. Arumugam, G.B. Bian, An intelligent learning approach for improving ECG signal classification and arrhythmia analysis, *Artif. Intell. Med.* 103 (2020) 101788.
- [9] J. Rahul, M. Sora, L.D. Sharma, et al., An improved cardiac arrhythmia classification using an RR interval-based approach, *Biocybern. Biomed. Eng.* 41 (2021) 656-666.
- [10] S. Nahak, A. Pathak, G. Saha, et al., Fragment-level classification of ECG arrhythmia using wavelet scattering transform, *Expert Syst. Appl.* 224 (2023) 120019.
- 570 [11] H.T. Shi, C.J. Qin, D.Y. Xiao, et al., Automated heartbeat classification based on deep neural network with multiple input layers, *Knowl-based Syst.* 188 (2020) 105036.
- [12] R.J. Ge, T.F. Shen, Y. Zhou, et al., Convolutional squeeze-and-excitation network for ECG arrhythmia detection, *Artif. Intell. Med.* 121 (2021) 102181.
- [13] P. Singh, A. Sharma, Interpretation and classification of arrhythmia using deep convolutional network, *IEEE Trans. Instrum. Meas.* 71 (2022) 2518512.

- 580 [14] C. Liu, S.B. Cheng, W.P. Ding, Spectral cross-domain neural network with soft-adaptive threshold spectral enhancement, *IEEE Trans. Neur. Net. Lear.* 36 (1) (2023) 692-703.
- [15] Y. Liu, C.J. Qin, Y.Y. Tian, et al., Lightweight optimization of deep learning models for accurate arrhythmia detection in clinical 12-lead ECG data, *IEEE Trans. Instrum. Meas.* 73 (2024) 2527717.
- [16] Y.H. Cheng, D. Li, D.D. Wang, et al., Multi-label arrhythmia classification using 12-lead ECG based on lead feature guide network, *Eng. Appl. Artif. Intel.* 129 (2024) 107599.
- 590 [17] H.X. Yang, Y.F. Zhao, X.W. Yang, PearsonUnet: ECG classification model integrating Pearson related beats, *IEEE Trans. Instrum. Meas.* 74 (2025) 2508011.
- [18] F. Liu, C.Y. Liu, L. Zhao, et al., An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection, *Journal of Medical Imaging and Health Informatics*, 8 (7) (2018) 1368-1373.
- [19] G.B. Moody, R.G. Mark, The impact of the MIT/BIH arrhythmia database, *IEEE Eng. Med. Biol. Mag.* 20 (3) (2001) 45-50.
- [20] X.S. Jia, P.H. Wu, L. Chen, et al., Heterogeneous driving graph transformer for multi-agent trajectory prediction via scene encoding, *IEEE Trans. Pattern. Anal.* 45 (11) (2023) 13860-13875.
- 600 [21] Z.K. Zhou, J.P. Wang, Y.H. Li, et al., Query-centric trajectory prediction, in: *Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR)*, 2023.
- [22] Z.K. Zhou, L.Y. Ye, J.P. Wang, et al., Hivt: hierarchical vector transformer for multi-agent motion prediction, in: *Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR)*, 2022.

- [23] Y.Y. Zhu, D. Luan, S.J. Shen. Biff: bi-level future fusion with polyline-based coordinate for interactive trajectory prediction, in: Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR), 2023.
- 610 [24] N. Nayakanti, R.A. Rfou, A. Zhou, et al., Wayformer: motion forecasting via simple and efficient attention networks, in: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), 2023.
- [25] X.S. Wang, T. Su, F. Da, et al., Prophnet: efficient agent-centric motion forecasting with anchor-informed proposals, in: Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR), 2023.
- [26] J.Y. Gao, C. Sun, H. Zhao, et al., Vectornet: encoding hd maps and agent dynamics from vectorized representation, in: Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR), 2020.
- 620 [27] M. Amrouche, D.M. Stipanovic, A formal characterization of activation functions in deep neural networks, *IEEE Trans. Neur. Net. Lear.* 35 (2) (2024) 2153-2166.
- [28] B.D. Weerdt, Y.C. Eldar, N. Deligiannis, et al., Deep unfolding transformers for sparse recovery of video, *IEEE Trans. Signal Proces.* 72 (2024) 1782-1796.
- [29] J. Kafkova, B. Babusiak, R. Pirnik, et al., Seat to beat: novel capacitive ECG integration for in-car cardiovascular measurement, *Measurement* 240 (2025) 115528.
- 630 [30] A. Petroni, F. Cuomo, G. Scarano, et al., MUSE: multi-lead sub-beat ECG for remote AI based atrial fibrillation detection, *J. Netw. Comput. Appl.* 212 (2023) 103544.
- [31] K.M. He, X.Y. Zhang, S.Q. Ren, et al., Deep residual learning for image recognition, in: Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR), 2016.

- [32] K.J. Dong, L. Zhao, C.R. Zou, et al., A novel ECG-derived respiration method combining frequency-domain feature and interacting multiple model smoother, *IEEE Trans. Biomed. Eng.* 70 (3) (2023) 888-898.
- [33] X.D. Peng, H.Q. Zhu, X. Zhou, et al., ECG signals segmentation using deep spatiotemporal feature fusion U-net for QRS complexes and R-peak detection, *IEEE Trans. Instrum. Meas.* 72 (2023) 2506212.
- 640 [34] G.R. Shaw, P. Savard, On the detection of QRS variations in the ECG, *IEEE Trans. Biomed. Eng.* 42 (7) (2002) 736-741.
- [35] L. Chen, Z.H. Jiang, J. Barker, et al., ECGVEDNET: a variational encoder-decoder network for ECG delineation in morphology variant ECGs, *IEEE Trans. Biomed. Eng.* 71 (7) (2024) 2143-2153.
- [36] B. Zhou, A. Khosla, A. Lapedriza, et al., Learning deep features for discriminative localization, in: *Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [37] R.R. Selvaraju, M. Cogswell, A. Das, et al., Grad-CAM: visual explanations from deep networks via gradient-based localization, in: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017.
- 650 [38] A. Chattopadhyay, A. Sarkar, P. Howlader, et al., Grad-CAM++: generalized gradient-based visual explanations for deep convolutional networks, in: *Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2018.
- [39] Q. Li, K. Li, X.Y. Gao, et al., Anomaly detection based on temporal attention network with adaptive threshold adjustment for electrical submersible pump, *IEEE Trans. Instrum. Meas.* 73 (2024) 3526214.
- [40] Y. Huang, Z.Y. Lan, L.J. Cao, et al., FocSAM: delving deeply into focused objects in segmenting anything, in: *Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR)*, 2024.
- 660

- [41] B.S. Tchinda, D. Tchiotsop, R. Tchinda, Classification of cardiac arrhythmias using ECG shape-adapted wavelets and BiLSTM network, *Measurement* 257 (2026) 118832.