Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A density-adaptive spatiotemporal interaction network with Motif Matrix for premature contraction detection

Jibin Wang

wangjibin0503@126.com

Bengbu Medical College

Bo Shi

Bengbu Medical College

Huixiang Wen

Bengbu Medical College

Haoyi Wang

Bengbu Medical College

Research Article

Keywords: Electrocardiogram classification, Motif matrix, Interaction information

Posted Date: September 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10381435/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A density-adaptive spatiotemporal interaction network with Motif Matrix for premature contraction detection

Jibin Wang[a,b,*], Bo Shi[a], Huixiang Wen[a], Haoyi Wang[a]

[a] *School of Medical Imaging, Bengbu Medical University, Bengbu 233000, China*
[b] *School of Mathematics, Tianjin University, Tianjin 300354, China*

Abstract

Premature atrial contraction (PAC) and premature ventricular contraction (PVC) are two predominant subtypes of cardiac arrhythmias characterized by ectopic electrical impulses. Conventional diagnostic approaches for PAC and PVC detection primarily rely on manual visual inspection of electrocardiograms, which is inherently labor-intensive, time-consuming and prone to diagnostic variability due to subjective interpretation. To this end, we develop a novel Density-Adaptive model based on Motif Matrix (NDAMM) that contains Adaptive Spatial Interaction block and Adaptive Temporal Interaction block for automated PAC and PVC classification. Specifically, a multi-scale feature fusion strategy is utilized to adaptively select an optimal scope of neighbors participating in information interaction in spatial dimension and an atrous pyramidal pooling mechanism is exploited to generate the interaction probability for each agent in temporal dimension. The effectiveness of our NDAMM is evaluated across the two public databases. The results achieve stable improvements over the existing approaches and strike a trade-off between performance and complexity. Visualization analysis further confirms the interpretability thus exhibiting application potential in several areas such as epilepsy prediction.

Keywords: Electrocardiogram classification, Motif matrix, Interaction information

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

1. Introduction

Premature cardiac contractions, resulting from premature electrical impulses originating at ectopic pacemaker sites, represent the most prevalent form of cardiac arrhythmia in clinical practice [1, 2]. These ectopic beats mainly manifest as two distinct types: premature atrial contraction (PAC) originating from the atria and premature ventricular contraction (PVC) arising from ventricular tissue. Clinically significant arrhythmias including ventricular tachycardia and sudden cardiac death have been associated with these premature beats, particularly in severe cases. Current diagnostic protocols predominantly rely on manual visual inspection of electrocardiograms (ECG) by clinicians-an approach that is inherently labor-intensive, time-consuming and prone to diagnostic variability [2]. These limitations emphasize the critical requirement for developing automated detection systems that demonstrate clinical generalizability, computational intelligence and operational efficiency to boost diagnostic accuracy and patient outcomes.

Machine learning (ML) approaches have demonstrated substantial progress in automated ECG analysis, exhibiting exceptional discriminative capacity for various cardiac arrhythmias. The design of feature extraction systems remains a critical determinant of algorithmic performance with established methodologies spanning statistical analyses, wavelet transformations and information-theoretic measures [3]. Nevertheless, the conventional ML implementations for ECG signal analysis present three persistent technical challenges: inherent limitations in algorithmic robustness under clinical variability, susceptibility to high-frequency overfitting issue in non-stationary physiological signals and the absence of standardized feature selection protocols across heterogeneous datasets [4, 5]. These fundamental constraints significantly impede the clinical translation of conventional ML strategies, particularly in real-world diagnostic applications requiring consistent performance across diverse patient populations.

Deep learning models (DLMs) have emerged as promising alternatives to conventional ML approaches, demonstrating superior generalizability and repre-

sentational capacity across several domains. This paradigm shift has redirected research focus from manual feature engineering to sophisticated architectural design, particularly through convolutional neural networks (CNN) that leverage inherent self-learning capabilities. DLMs enable automated hierarchical feature extraction while maintaining robust performance across heterogeneous data distributions. Contemporary architectures predominantly adopt either CNN-based frameworks (e.g., ConvNet [6]) or transformer-based designs (e.g.,Vision Transformer [7], Swin Transformer [8]). However, the existing architectures often fail to explicitly incorporate domain-specific knowledge regarding spatial-temporal information interaction inherent in ECGs [9].

Based on this, we develop a novel Density-Adaptive network based on Motif Matrix (NDAMM) model that contains Adaptive Spatial Interaction (ASI) block and Adaptive Temporal Interaction (ATI) block. Our NDAMM is capable of distilling interaction information from the perspectives of Spatial and Temporal dimensions. Specifically, a multi-scale feature fusion strategy is exploited to adaptively select an optimal scope of neighbors participating into information interaction in spatial dimension; temporal interaction features are extracted and also an atrous pyramidal pooling mechanism is leveraged to generate the interaction probability for each agent in temporal dimension. Further, we conduct a series of ablation studies to confirm the effectiveness of NDAMM for PAC and PVC detection across both China Physiological Signal Challenge 2018 database (CPSCDB) and MIT-BIH arrhythmia database (MITDB).

Our contributions include several aspects:

- **Novel Spatiotemporal Interaction Modeling** We develop NDAMM, a density-adaptive framework integrating Adaptive Spatial and Temporal Interaction blocks, to explicitly model information interactions across spatial and temporal dimensions for cardiac arrhythmia detection.

- **Superior Performance with Lower Complexity** Extensive validation across CPSCDB and MITDB demonstrates better accuracy in PAC/PVC detection achieving optimal performance-complexity tradeoffs than several

existing strategies.

- **Enhanced Interpretability and Potential** Visualization analysis confirms the interpretability and exhibits potential for broader applications such as epilepsy prediction.

2. Recent work

In this section, we review the research on PAC and PVC detection in recent years and they are mainly divided into two categories: machine learning-Based methods (MLM) and deep learning-Based models (DLM).

- **Machine learning – Based methods**: Zhu et al. [10] introduced a pattern recognition mechanism with ECG features and support vector machine (SVM) classifier to discriminate various ECG types including PAC and PVC. Oliveira et al. [11] developed a methodology based on analytic hierarchy process and an SVM classifier for the identification of PAC and PVC episodes. Sangaiah et al. [12] proposed an authentication system based on wavelet coefficient features and hidden Markov model (HMM) for PAC and PVC identification. Jha et al. [13] constructed an ML approach technique applying wavelet features of ECG fragments and an SVM classifier to distinguish PAC, PVC and normal sinus rhythm (NSR) signals. Rahul et al. [14] investigated a robust strategy to utilize RR intervals and statistical features with various classifier for the recognition of PAC and PVC. Besides, Nahak et al. [15] exploited an ML approach that integrates wavelet scattering features and an SVM classifier for the analysis of PAC and PVC fragments. Despite demonstrating feasibility, conventional MLM exhibit several inherent limitations. Their reliance on handcrafted feature extraction introduces subjectivity, compromises robustness under clinical variability and hinders generalizability across diverse datasets. Thus, their clinical translation for reliable PAC/PVC detection remains challenging, motivating the adoption of DLM capable of automated, hierarchical feature learning.

- **Deep learning – Based models**: Bouny et al. [16] come up with a multi-level wavelet convolutional neural networks (ML-WCNN) for PAC and PVC detection. Ge et al. [17] introduced an Squeeze-and-Excitation (SE) model for automatic detection of several arrhythmias. Zhang et al. [18] displayed an attention U-net framework for the identification of PAC and PVC segments. Zahid et al. [19] proposed a DLM approach based on self-operational neural network to discriminate PAC and PVC signals. Park et al. [20] investigated a PAC and PVC recognition scheme using fully convolutional network (FCN) and long short-term memory (LSTM). Liu et al. [21] exploited an optimized module replacement method with pure CNN-based architecture for the recognition of PAC and PVC. Cheng et al. [22] constructed a lead feature guide network (LFG-Net) for automatic detection of PAC and PVC fragments. Sarankumar et al. [23] implemented a DLM detection strategy based on bidirectional gated recurrent unit with auto encoder to distinguish PAC and PVC episodes. Yang et al. [24] proposed an ECG classification framework that integrates parallel-based U-net and Transformer with Pearson related beats. While the existing DLMs offer superior representational capacity over ML methods, a critical gap persists: existing architectures often fail to explicitly incorporate domain-specific knowledge regarding spatial-temporal information interaction inherent in ECGs. To address this, we propose NDAMM, explicitly designed to distill adaptive interaction features across both spatial and temporal dimensions.

3. Methods and materials

3.1. Database

To validate the NDAMM module's generalizability, we conducted comprehensive evaluations using two publicly available ECG datasets:

- MIT-BIH Arrhythmia Database (**MITDB**) [25]: contains 48 dual-channel recordings from 47 subjects, each spanning 30 min at 360 Hz sampling rate.

- China Physiological Signal Challenge 2018 Database (**CPSCDB**) [26]: comprises 6,877 12-lead recordings (6-60 s duration, 500 Hz sampling rate) collected across 11 clinical centers.

For robust model training, we implemented a systematic preprocessing pipeline:

- **Resampling**: downsampled all signals to 250 Hz with anti-aliasing filters to ensure Nyquist compliance.
- **Denoising**: applied wavelet thresholding and bandpass filtering to suppress baseline wander and high-frequency noise.
- **Segmentation**: divided continuous recordings into 2 s clinically interpretable fragments.
- **Normalization**: performed channel-wise Z-score standardization to mitigate outlier effects.

The dataset composition and representative waveforms of PAC, PVC and NSR are summarized in Table 1.

Table 1: The particulars of ECG fragments in this work.

Database	ECG type	Quantity
China Physiological Signal Challenge 2018 database (CPSCDB)	NSR	9384
	PAC	6520
	PVC	7213
MIT-BIH arrhythmia database (MITDB)	NSR	6257
	PAC	3564
	PVC	4083

3.2. NDAMM

In an ECG signal episode with N agents as sample points S_i for $i \in \Lambda = \{1, ..., N\}$, the previous point trajectory X of S_i across the time steps $t \in [-t_-, 0]$

is determined as $X_i = (q_i^{-t_-}, ..., q_i^0)$ and the next point trajectory Y over the time steps $t \in [0, t_+]$ is determined as $Y_i = (q_i^0 ..., q_i^{t_+})$. We aim to predict the distribution of next point trajectories $P(Y|X, Z)$ with previous point trajectory X and feature representation Z. For the next behavior of each agent, we use a series of latent variables of agent i as $\Psi = \{\Psi_i\}_{i \in \Lambda}$ and the next point trajectory distribution is described:

$$\begin{cases} P(Y|X,Z) = \sum_i [P(Y|\Psi_i, X, Z) \cdot P(\Psi_i|X, Z)] \\ \qquad\qquad \triangleq \sum_i P_D \cdot P_\theta \\ with\ [\Psi_i = \Psi_{s_i} \cup \Psi_{t_i}] \wedge [\Psi_{s_i} \cap \Psi_{t_i} \neq \emptyset] \end{cases} \tag{1}$$

where Ψ_{s_i} represents spatial features and Ψ_{t_i} represents temporal features. We

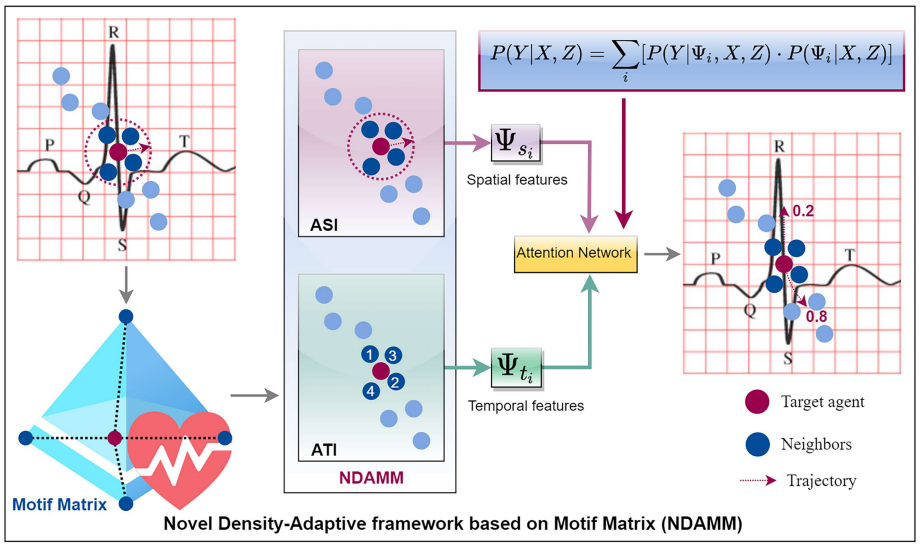


Fig. 1. Architecture overview of novel Density-Adaptive framework based on Motif Matrix (NDAMM). This models the interaction procedure between target agent and its neighbors with density adaptation, ASI describes the range for multi-scale neighbors selection and ATI describes the dynamic interaction analysis (where numbers are interaction probabilities). ECG features from ASI and ATI are combined by attention architecture to discriminate diverse trajectories (where number represents generation probability).

use Conditional Variational AutoEncoder (CVAE) architecture for accuracy and

generalization [27]. With the help of Motif Matrix [28], our NDAMM consists of both Adaptive Spatial Interaction (ASI) block and Adaptive Temporal Interaction (ATI) block in Fig.1. Each agent i is sequentially treated as the target agent in the scenes and so we focus the interaction between sample points S_i and its neighbors $\{S_j\}_{j\neq i}^{N}$.

3.2.1. Adaptive Spatial Interaction

In an ASI block in Fig.2, it is essential to determine which neighbors interact with the target agent within the scopes. To this end, we capture fusion features ξ from both Motif Matrix and agent, simultaneously integrating them with each agent feature ψ_i. Also, we obtain the Motif Matrix $M_{s,t}$ in the spatial dimension $M_{s,*}$. Further, a motif matrix for each radius by erasing the temporal dimension can be obtained:

$$\widetilde{M_s} = \sum_t M_{s,t}(s = s_1, ..., S) \tag{2}$$

where s denotes the scope for selecting neighbors with s_1 to S, thus the motif matrices in spatial dimension are obtained as $\widetilde{M_{s_1}}, ..., \widetilde{M_S}$.

To acquire the primary features $\psi_s^{\widetilde{M}}$ from motif matrix $\widetilde{M_s}$, we make use of Atrous Spatial Pyramid Pooling (ASPP) network [29] to aggregate the features from each $\widetilde{M_s}$ into a unified dimension:

$$\psi_s^{\widetilde{M}} = LN\Big\{FC\big[ASPP(\widetilde{M_s})\big]\Big\} \tag{3}$$

where FC is Fully-Connected layer and LN is Layer Normalization. To fuse features from both agents and motif matrices, a multi-head attention mechanism is utilized within each agent i by Queries Q concatenating agents feature and Keys K and Values V capturing representation from different motif matrices as follows:

$$\begin{aligned} &MultiHead(Q_i, K_s, V_s) \\ &= \Big[W^Q(\psi_i; \psi_G), W^K\psi_s^{\widetilde{M}}, W^V\psi_s^{\widetilde{M}}\Big] \end{aligned} \tag{4}$$

where $i \in \Lambda$, W^Q, W^K, W^V are learnable weight matrices, semicolon indicates concatenation and comma indicates components of a vector.

Further, we can obtain the features of both each agent ψ_i and global feature ψ_G in Eq.(4):

$$\begin{cases} \psi_i = GRU\big[MLP(X_i)\big] \\ \psi_G = MLP\big[PointNet^{+}(X_1;...;X_N)\big] \end{cases} \tag{5}$$

where GRU and MLP indicate Gated Recurrent Unit and Multi-Layer Perceptron respectively. $PointNet^{+}$ [30] is applied to remain consistent dimensions for agents within diverse scenes. To alleviate the effect of large dot product values, scaled dot-product attention is utilized as follows:

$$\begin{cases} \xi_{i,s} = MultiHead(Q_i, K_s, V_s) \\ \qquad\quad = (head_1;...;head_h)W^O \\ with\ head_k = Attention(Q_i W_k^A, K_s W_k^A, V_s W_k^A) \end{cases} \tag{6}$$

To alleviate the information loss issue of motif matrix feature owing to ASPP in Eq.(5), three statistics maximum, minimum and mean are leveraged to ξ_i derived from $\{\xi_{i,s_1},...,\xi_{i,S}\}$ and we integrate these original features with statistical features via:

$$\xi_i^{\widetilde{M}} = MLP(\xi_{i,s_1};...;\xi_{i,S};\xi_{i,max};\xi_{i,min};\xi_{i,mea}) \tag{7}$$

The decomposed features of agent i in the spatial dimension can be determined:

$$\mathbf{\Psi}_{s_i} = \xi_i^{\widetilde{M}} \odot \psi_i \tag{8}$$

where $\odot$ indicates Hadamard product

3.2.2. Adaptive Temporal Interaction

To discriminate the neighbors that produce the significant influences on the target agent from ASI block, an optimal coefficient η is derived via feature fusion to aggregate matrices $\widehat{M}_*$ as $\sum \eta \widehat{M}_*$. In ATI block, we aim to acquire temporal features from motif matrix $M_{s,t}$ and $\widehat{M}_t$ at each step with a strategy analogous to that in Eq.(2) is described:

$$\widehat{M}_t = \sum_s M_{s,t}(t = t_1, ..., T) \tag{9}$$

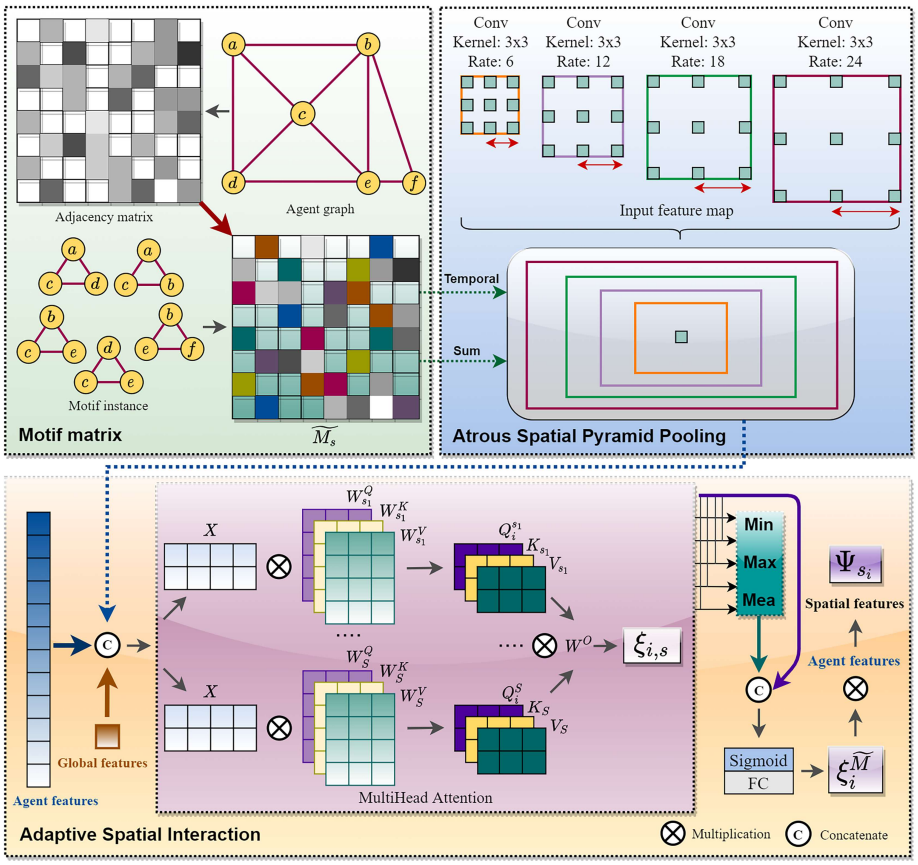


Fig. 2. Architecture overview of Adaptive Spatial Interaction.

where T represents the horizon of trajectory sequence with t_1 of initial step. We further combine the features of $\widehat{M}_*$:

$$\psi^{\widehat{M}} = LN\Big\{FC\big[ASPP(\widehat{M_{t_1}}); ...; ASPP(\widehat{M_T})\big]\Big\} \tag{10}$$

we aggregate $\psi^{\widehat{M}}$ and statistical functions as Eq.(7) to alleviate information loss issue as follows:

$$\eta^{\widehat{M}} = MLP(\psi^{\widehat{M}}; \psi^{\widehat{M}}_{max}; \psi^{\widehat{M}}_{min}; \psi^{\widehat{M}}_{mea}) \tag{11}$$

both global trajectory feature ψ_G and temporal motif features $\eta^{\widehat{M}}$ determine the weight parameters of temporal motif matrix, which are acquired via ROI-wise Reverse Reweighting Network [31]:

$$Reweighting(\psi_G; \eta^{\widehat{M}}) \rightarrow \eta^{\widehat{M}} = (\eta_1, ..., \eta_T) \tag{12}$$

The decomposed features of agent i in the temporal dimension are determined in Fig.3:

$$\Psi_{t_i} = \sum_{t_1}^{T} \eta_t \widehat{M}_t \tag{13}$$

In brief, the feature outputs of NDAMM through concatenating the decomposed spatial features Ψ_{s_i} and temporal features Ψ_{t_i} are indicated as $\Psi_i = [\Psi_{s_i}; \Psi_{t_i}]$. The specific algorithm pseudocode is shown in Algorithm 1.

3.3. Training the Loss

Our loss function consists of three components: **Classification loss** L_{CLS}, **Selection Binary Cross Entropy (BCE) Loss** L_{SEL} and **Kullback-Leibler divergence loss** L_{KLD}. For each sample point i, we quantify the discrepancy between the predicted class probability distribution $\hat{Y}_i$ (PAC, PVC, NSR) and the ground-truth label y_i using categorical cross-entropy:

$$L_{\mathrm{CLS}} = -\sum_i y_i \log\left(\mathrm{softmax}\left[\mathrm{MLP}(\Psi_i; \Psi_G; \Psi_L)\right]\right) \tag{14}$$

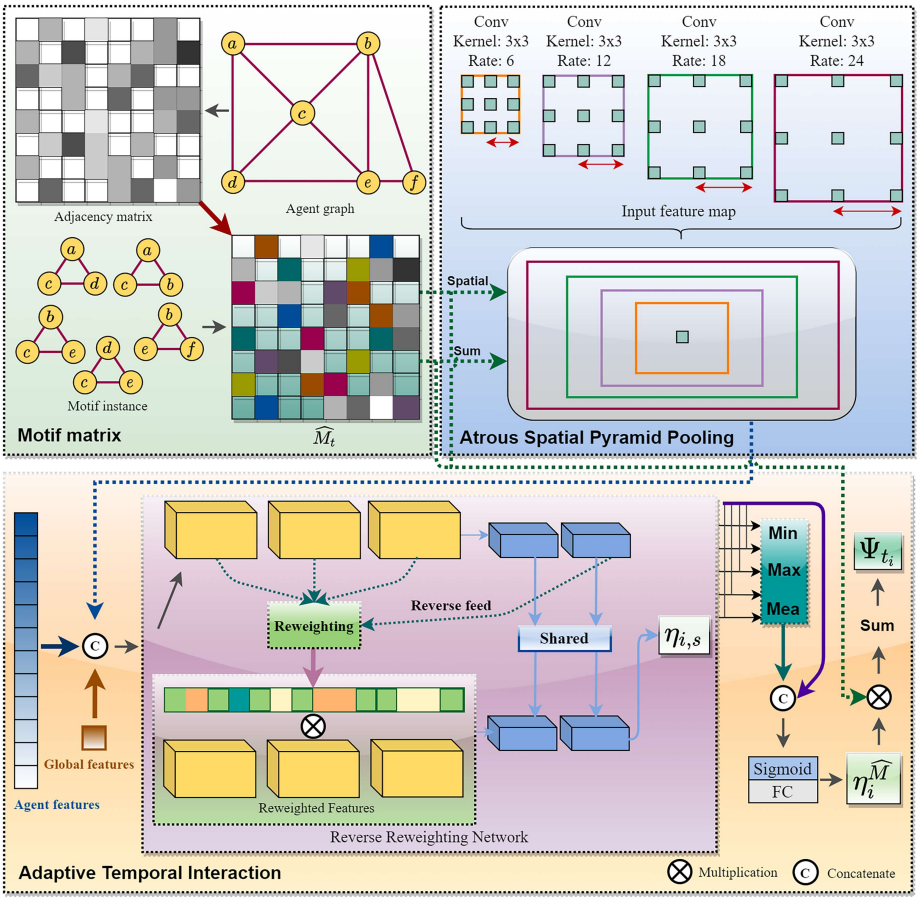


Fig. 3. Architecture overview of Adaptive Temporal Interaction.

Algorithm 1 NDAMM Pseudocode

Require: - Sample points set $S = \{S_i\}$, $i \in \Lambda = \{1, \ldots, N\}$

- Historical trajectory $X_i = (q_i^{-t-}, \ldots, q_i^0)$

- Motif matrix $M_{s,t}$

Ensure: - Predicted next trajectory distribution $P(Y|X, Z)$

▷ **Spatial Interaction (ASI Block)**

for each spatial scope $s \in \{s_1, \ldots, S\}$ **do**

  Compute aggregated Motif matrix: $\widetilde{M}_s = \sum_t M_{s,t}$

  Extract spatial feature: $\psi_s^{\widetilde{M}} = \mathrm{LN}\{\mathrm{FC}[\mathrm{ASPP}(\widetilde{M}_s)]\}$

end for

for each sample point $i \in \Lambda$ **do**

  Local feature: $\psi_i = \mathrm{GRU}[\mathrm{MLP}(X_i)]$

  Global feature: $\psi_G = \mathrm{MLP}[\mathrm{PointNet}^+(X_1; \ldots; X_N)]$

  Multi-head attention fusion:

  $Q_i = W^Q[\psi_i; \psi_G]$

  $K_s = W^K \psi_s^{\widetilde{M}}, V_s = W^V \psi_s^{\widetilde{M}}$

  $\xi_{i,s} = \mathrm{MultiHead}(Q_i, K_s, V_s)$

  $\xi_i^{\widetilde{M}} = \mathrm{MLP}(\xi_{i,s_1}; \ldots; \xi_{i,S}; \max; \min; \mathrm{mean})$

  Spatial decomposed feature: $\Psi_{s_i} = \xi_i^{\widetilde{M}} \odot \psi_i$

end for

▷ **Temporal Interaction (ATI Block)**

for each timestep $t \in \{t_1, \ldots, T\}$ **do**

  Compute aggregated Motif matrix: $\widetilde{M}_t = \sum_s M_{s,t}$

end for

$\psi^{\widetilde{M}} = \mathrm{LN}\big\{\mathrm{FC}\{[\mathrm{ASPP}(\widetilde{M}_{t_1}); \ldots; \mathrm{ASPP}(\widetilde{M}_T)]\}\big\}$

$\eta^{\widetilde{M}} = \mathrm{MLP}(\psi^{\widetilde{M}}; \max; \min; \mathrm{mean})$

Temporal weight generation: $\eta = \mathrm{Reweighting}(\psi_G; \eta^{\widetilde{M}})$

for each sample point $i \in \Lambda$ **do**

  Temporal decomposed feature: $\Psi_{t_i} = \sum_{t=1}^{T} \eta_t \cdot \widetilde{M}_t$

end for

▷ **Feature Fusion & Output**

for each sample point $i \in \Lambda$ **do**

  Final feature representation: $\Psi_i = [\Psi_{s_i}; \Psi_{t_i}]$

  Trajectory distribution: $P(Y|X, Z) = \sum_i P_D \cdot P_\theta$

end for

where Ψ_L denotes lead-specific features. We further ensure alignment between latent interaction selections and ground-truth labels:

$$L_{\mathrm{SEL}} = \mathrm{BCE}\left\{g^m, \mathrm{softmax}\left[\mathrm{MLP}(\psi_i; \psi_G)\right]\right\} \tag{15}$$

where g^m is one-hot vector of the ground-truth label and ψ_i, ψ_G are agent/global features. The latent space by minimizing KL divergence between the posterior $q(\Psi_i|X, Z)$ and prior $p(\Psi_i)$ distributions are regularized:

$$L_{\mathrm{KLD}} = D_{\mathrm{KL}}\left(q(\Psi_i|X, Z) \parallel p(\Psi_i)\right) \tag{16}$$

The composite loss is a weighted sum:

$$\mathcal{L} = L_{\mathrm{CLS}} + \lambda_1 L_{\mathrm{SEL}} + \lambda_2 L_{\mathrm{KLD}}, \tag{17}$$

where λ_1, λ_2 are trade-off parameters. λ_2 increases progressively during training to enhance latent space stability.

3.4. Architecture details

To validate the efficacy of our proposed NDAMM, we developed NDAMM-ResNet50 by systematically integrating the module into the standard ResNet50 architecture. As displayed in Fig.4, this network architecture leverages five sequentially stacked residual building blocks, each containing four strategically positioned NDAMM modules. The final classification layer utilizes Softmax activation to generate probability distributions for three distinct cardiac arrhythmia categories: PAC, PVC and NSR. For model optimization, we implemented Adam optimizer with a learning rate of 0.001 and trained the network using mini-batches of 256 samples to ensure stable convergence.

4. Results

In the specific experiments, we conduct a series of ablation studies with a total of 23117 ECG fragments involving 9384 NSR, 6520 PAC and 7213 PVC on CPSCDB as well as 13904 ECG fragments involving 6257 NSR, 3564 PAC

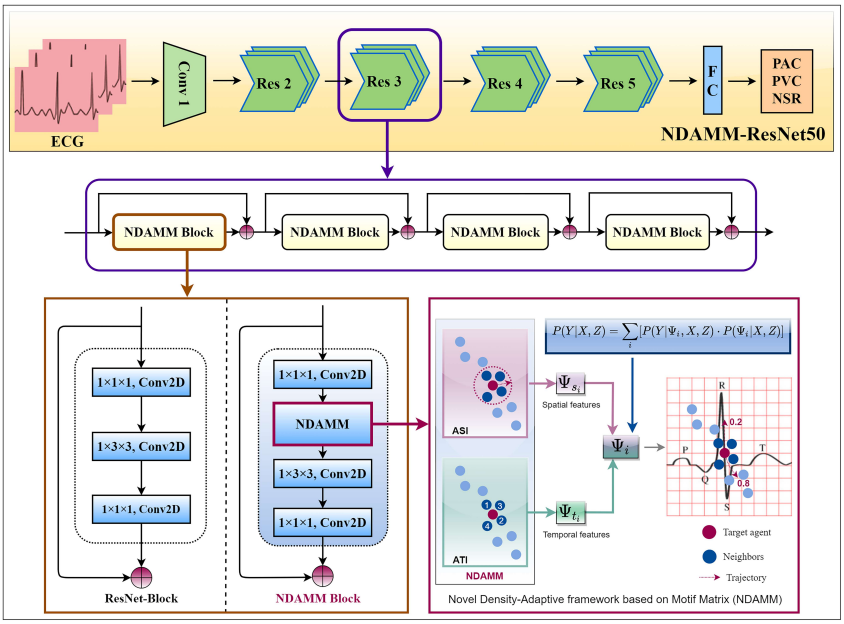


Fig. 4. Architecture overview of NDAMM-ResNet50.

and 4083 PVC on MITDB. To confirm the effectiveness of NDAMM, the existing Vision Transformer (ViT) [7], Swin Transformer (SwinT) [8], and DAMM [9] are utilized to construct ViT-ResNet50, SwinT-ResNet50 and DAMM-ResNet50 and these frameworks are constructed by the same compositions as NDAMM-ResNet50, except that our NDAMM is replaced with ViT, SwinT and DAMM, respectively. Also, accuracy (ACC), sensitivity (SEN) and specificity (SPF) are exploited to evaluate model performance.

4.1. Signal recognition

Table 2 demonstrates the recognition results of five proposed architectures with CPSCDB. 99.0% of ECG fragments are accurately classified as NSR fragments via NDAMM-ResNet50, 99.1% of ECG fragments are accurately sorted as PAC fragments, 1.0% of PVC fragments are falsely distinguished as other heartbeats. For naive ResNet50, ViT-ResNet50, SwinT-ResNet50 and DAMM-ResNet50, 98.2%, 98.7%, 98.7% and 99.0% of ECG fragments are precisely sorted as NSR fragments, respectively. 98.2%, 98.7%, 98.8% and 98.9% of ECG

fragments can be precisely distinguished as PAC episodes, respectively. 1.7%, 1.4%, 1.2% and 1.1% of PVC fragments are falsely distinguished as other heartbeats, respectively.

Table 2: Recognition performance results of the proposed architectures with CPSCDB.

Architecture	Category	ACC(%)	SEN(%)	SPF(%)
Naive ResNet50	NSR	98.8	98.2	99.0
	PAC	98.9	98.2	98.9
	PVC	98.8	98.3	98.9
ViT-ResNet50 [7]	NSR	99.2	98.7	99.2
	PAC	99.2	98.7	99.4
	PVC	99.1	98.6	99.3
SwinT-ResNet50 [8]	NSR	99.2	98.7	99.3
	PAC	99.2	98.8	99.4
	PVC	99.3	98.8	99.3
DAMM-ResNet50 [9]	NSR	99.3	99.0	99.4
	PAC	99.3	98.9	99.4
	PVC	99.4	98.9	99.3
NDAMM-ResNet50	NSR	99.5	99.0	99.4
	PAC	99.4	99.1	99.5
	PVC	99.4	99.0	99.5

Table 3 demonstrates the recognition results of five proposed architectures with MITDB. 98.7% of ECG fragments are accurately classified as NSR fragments via NDAMM-ResNet50, 98.8% of ECG fragments are accurately sorted as PAC fragments, 1.2% of PVC fragments are falsely distinguished as other heartbeats. For naive ResNet50, ViT-ResNet50, SwinT-ResNet50 and DAMM-ResNet50, 98.0%, 98.4%, 98.5% and 98.6% of ECG fragments are precisely sorted as NSR fragments, respectively. 97.8%, 98.3%, 98.5% and 98.7% of ECG

fragments can be precisely distinguished as PAC episodes, respectively. 2.1%, 1.7%, 1.6% and 1.4% of PVC fragments are falsely distinguished as other heartbeats, respectively.

Table 3: Recognition performance results of the proposed architectures with MITDB.

Architecture	Category	ACC(%)	SEN(%)	SPF(%)
Naive ResNet50	NSR	98.5	98.0	98.9
	PAC	98.4	97.8	98.8
	PVC	98.7	97.9	98.7
ViT-ResNet50 [7]	NSR	98.9	98.4	99.2
	PAC	99.0	98.3	99.3
	PVC	98.9	98.3	99.2
SwinT-ResNet50 [8]	NSR	98.9	98.5	99.1
	PAC	99.0	98.5	99.2
	PVC	99.0	98.4	99.3
DAMM-ResNet50 [9]	NSR	99.1	98.6	99.3
	PAC	99.0	98.7	99.2
	PVC	99.1	98.6	99.3
NDAMM-ResNet50	NSR	99.1	98.7	99.4
	PAC	99.2	98.8	99.3
	PVC	99.2	98.8	99.4

Table 2 and Table 3 demonstrate that NDAMM-ResNet50 attains the best recognition performance than other control group models with both databases. Performance measurements are 0.1% to 0.9% better than those of control group models authenticating the validity of NDAMM and more prominent recognition ability than standard DAMM, ViT and SwinT for PAC and PVC. NDAMM-ResNet50 still achieves a better performance than DAMM-ResNet50. This result demonstrates the high efficiency of NDAMM and we assume that the main

reason for performance improvement is the application of feature $Mean$ and ASPP. Since the operations control the resolution of output feature maps and also increase the receptive field of convolution kernel effectively compared to the original DAMM [29].

4.2. Model convergence

The convergence behavior of NDAMM-ResNet50, as depicted in Fig.5 (CPSCDB) and Fig.6 (MITDB), demonstrates remarkable stability and efficiency across both datasets. The model achieves convergence within 23-24 epochs for CPSCDB and 22-23 epochs for MITDB maintaining lower oscillation amplitudes during optimization. This accelerated convergence stems from NDAMM's intrinsic architectural advantages: 1) Adaptive Spatial Interaction block dynamically regulates feature aggregation scope through multi-scale fusion, preventing gradient instability; 2) Adaptive Temporal Interaction block employs pyramidal pooling to maintain temporal feature consistency during backpropagation; 3) Motif matrix formulation provides mathematical guarantees for convex optimization landscapes under the specified parameter constraints. Such rapid yet stable convergence validates NDAMM's robustness against distributional shifts in clinical ECG. The smooth decay of training and validation loss curves further confirms effective generalization without over-fitting risks.

4.3. Detail analysis

Table 4 and Table 5 manifest overall identification performance for NDAMM-ResNet50 with different feature aggregation strategies in Eq.(3) with both databases. NDAMM-ResNet50 achieves the best detection performance when Atrous Spatial Pyramid Pooling strategy is exploited. We speculate that compared with Spatial Pyramid Pooling (SPP) and Depthwise Separable Pyramidal Pooling (DSPP), a prominent reason for performance improvement is that this strategy can increase the receptive field of convolution kernel effectively without sacrificing complexity. Fig.7 shows the change trend of performance measurements

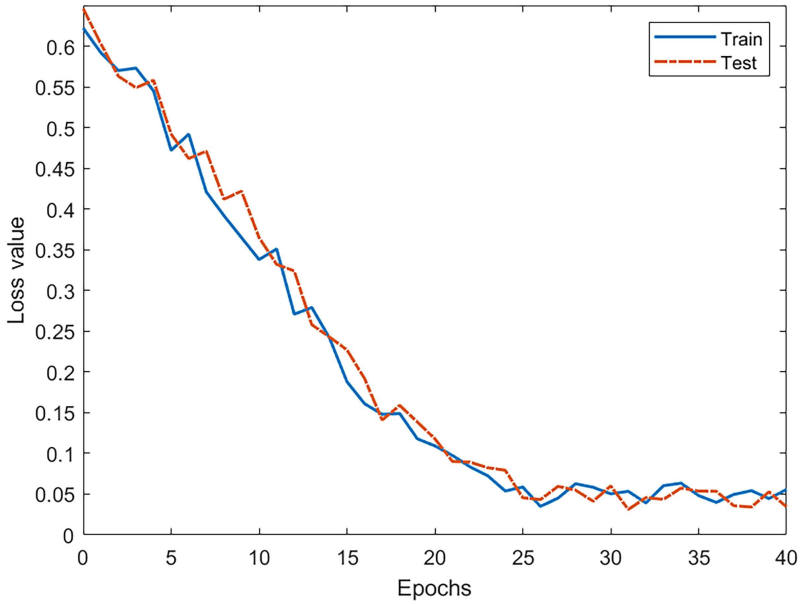


Fig. 5. Train curve tendency for NDAMM-ResNet50 with CPSCDB.

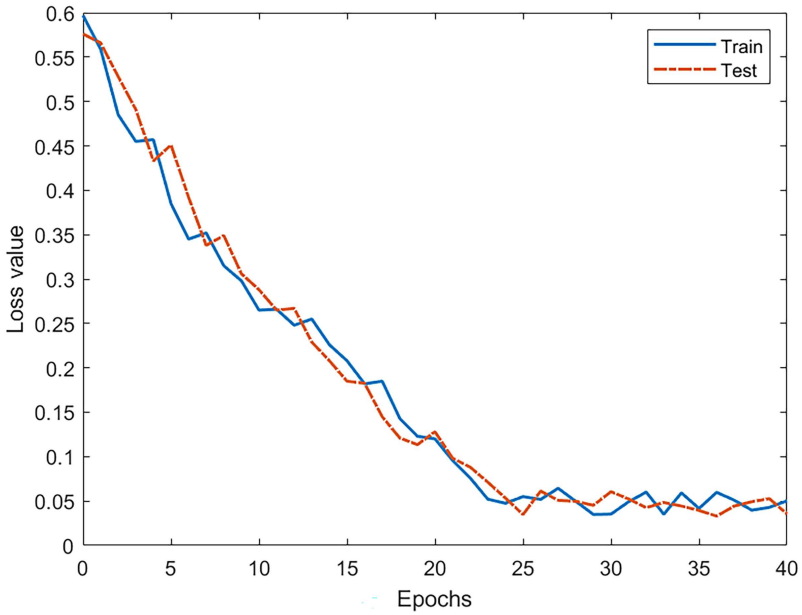


Fig. 6. Train curve tendency for NDAMM-ResNet50 with MITDB.

Table 4: Overall recognition performances for NDAMM-ResNet50 with different feature aggregation strategies in Eq.(3) with CPSCDB.

Feature aggregation strategies	ACC(%)	SEN(%)	SPF(%)
Naive version	98.3	98.2	98.4
Spatial Pyramid Pooling [32]	99.0	98.9	99.0
Depthwise Separable Pyramidal Pooling [33]	99.1	99.0	99.1
Atrous Spatial Pyramid Pooling (Ours)	99.4	99.0	99.5

Table 5: Overall recognition performances for NDAMM-ResNet50 with different feature aggregation strategies in Eq.(3) with MITDB.

Feature aggregation strategies	ACC(%)	SEN(%)	SPF(%)
Naive version	98.0	98.0	97.9
Spatial Pyramid Pooling [32]	98.7	98.6	98.8
Depthwise separable pyramidal pooling [33]	98.8	98.7	98.7
Atrous Spatial Pyramid Pooling (Ours)	99.2	98.8	99.4

for NDAMM-ResNet50 with different feature aggregation strategies across both databases attesting the conclusion again.

Table 6 and Table 7 manifest overall recognition performance for NDAMM-ResNet50 with different statistical features in Eq.(7) on two databases. NDAMM-ResNet50 achieves the best detection performance when $Max; Min; Mea$ strategy is integrated into Eq.(7). The elemental reason is that Mean strategy can maintain the stability of feature changes than other statistical features such as Standard Deviation. Fig.8 illustrates the change trend of performance measurements for NDAMM-ResNet50 with different statistical features on both databases attesting the conclusion again.

Table 6: Overall recognition performances for NDAMM-ResNet50 with different statistical features in Eq.(7) with CPSCDB. Here Median is abbreviated as Med; Standard Deviation as SD; Mean as Mea.

Statistical features	ACC(%)	SEN(%)	SPF(%)
Naive version	98.2	98.3	98.2
$Max; Min; SD$ [34]	98.6	98.5	98.7
$Max; Min; Med$ [9]	99.0	98.9	99.0
$Max; Min; Mea$ (Ours)	99.4	99.0	99.5

5. Discussion

5.1. Complexity analysis

To manifest inference efficiency of NDAMM, we reproduced many existing classical strategies and executed the same pre-processing steps and experimental configuration with CPSCDB. The number of hyper-parameters and floating-point operations (FLOPs) of network architectures are calculated to assess their time and space complexity in Table 8 and Fig.9. Due to wavelet transform, dense connection in CNN and self-attention mechanism in Transformer, the time

Table 7: Overall recognition performances for NDAMM-ResNet50 with different statistical features in Eq.(7) with MITDB.

Statistical features	ACC(%)	SEN(%)	SPF(%)
Naive version	98.0	98.0	97.9
$Max; Min; SD$ [34]	98.4	98.3	98.4
$Max; Min; Med$ [9]	98.8	98.7	98.9
$Max; Min; Mea$ (Ours)	99.2	98.8	99.4

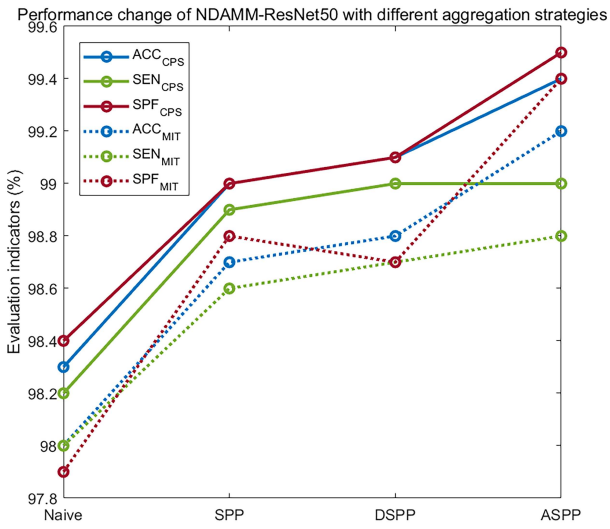


Fig. 7. Performance measurements change for NDAMM-ResNet50 with different feature aggregation strategies with both databases.

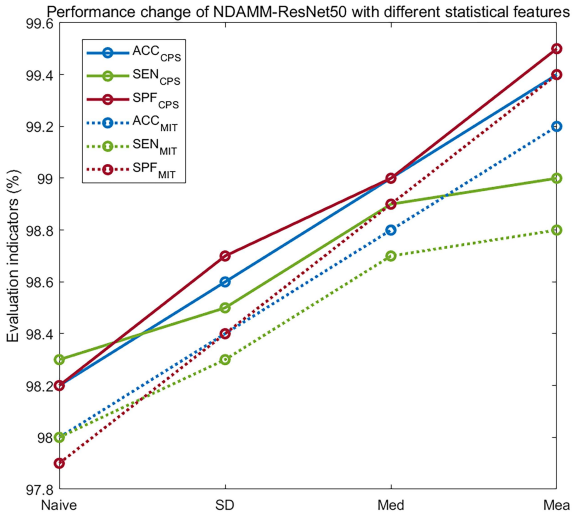


Fig. 8. Performance measurements change for NDAMM-ResNet50 with different statistical features with both databases.

and space complexity of NDAMM are reduced by 19.2% and 49.0% as well as 48.1% and 60.0% than ML-WCNN [16] and CNN-Transformer [24] architecture, respectively. The result indicates that our NDAMM has high inference efficiency and promising potential for mobile devices [36, 37]. There are similar conclusions in MITDB.

5.2. Visualization

We conducted visualization analysis of our NDAMM with gradient-weighted class activation mapping (Grad-CAM) for PAC and PVC detection in Fig.10.

Fig.10 indicates that the heatmap activation peak (red) is precisely localized at the abnormal P-wave region ($t = 0.58\,s$), validating ASI block's capability to capture atrial ectopic pacemakers through multi-scale neighborhood selection in Eq.(2). Atrous pyramidal pooling in Eq.(3) expands the local receptive field enabling the model to focus on P-wave morphological variations, which aligns with clinical PAC diagnostic criteria. The activation zone spans the entire wide QRS complex ($t = 0.7 \sim 0.86\,s$), reflecting the temporal interaction mechanism of

Table 8: Complexity comparisons of our proposed NDAMM with many existing classical strategies across CPSCDB.

Frameworks	Parameters (M)	FLOPs (GMac)	ACC(%)
ResNet [35]	1.10	0.89	98.2
ML-WCNN[16]	2.76	2.43	99.7
U-Net [18]	2.92	2.57	94.8
Transformer [24]	4.30	3.10	99.6
DAMM [9]	2.89	1.65	99.3
$NDAMM(Ours)$	2.23	1.24	99.4

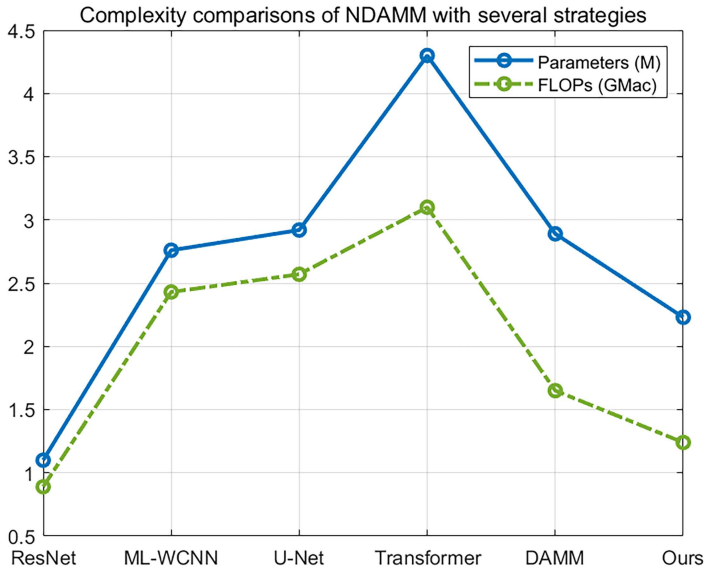


Fig. 9. Complexity evaluation of NDAMM with existing strategies.

ATI block. Reweighting network dynamically allocates weights, explaining the boundary enhancement phenomenon at QRS onset/offset in the heatmap. This characteristic demonstrates NDAMM's superiority over traditional CNN's local feature limitations. The visualizations demonstrate the model's spatiotemporal adaptive capabilities. The distinct heatmap distribution patterns (focal activation for PAC vs. band-like coverage for PVC) reflect the pathological essence of different arrhythmias demonstrating model interpretability.

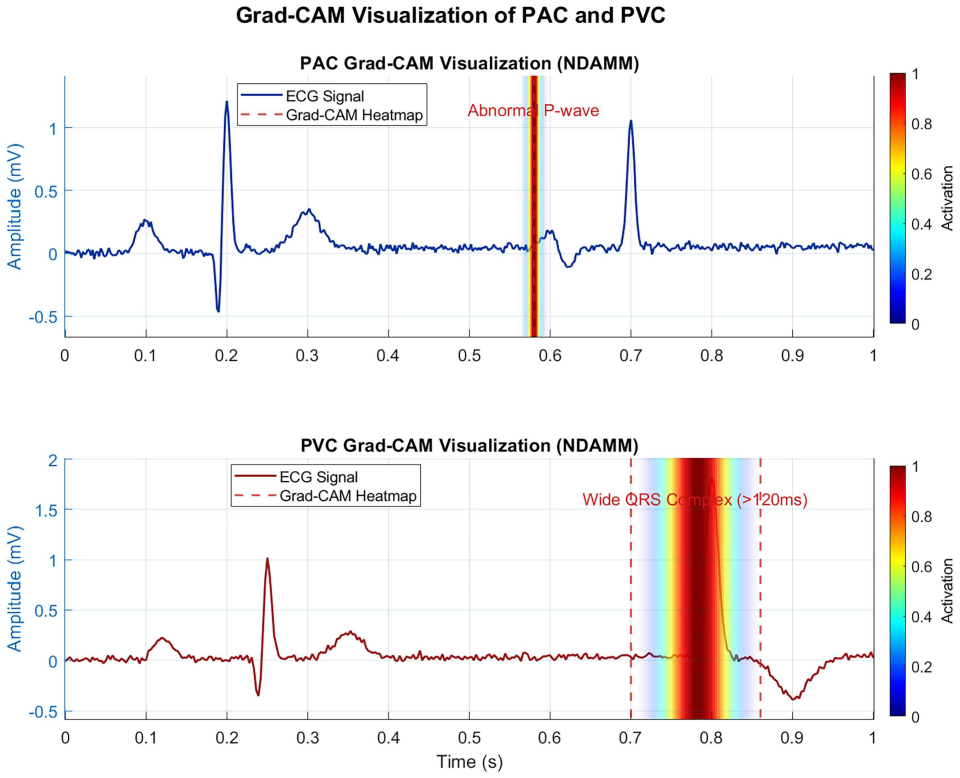


Fig. 10. Visualization analysis of NDAMM with gradient-weighted class activation mapping.

5.3. Performance comparison

Table 9 shows the performance comparison of NDAMM-ResNet50 with several existing schemes for PAC and PVC detection with both databases.

The researches on PAC and PVC identification are mainly divided into

Table 9: Performance comparison of NDAMM-ResNet50 with several existing schemes for PAC and PVC detection with both databases.

Researchers	Year	Approach	Data set	Performance(%)
Zhu et al.[10]	2018	ECG features+SVM	MITDB	ACC=97.80
Oliveira et al.[11]	2019	Morphological features+SVM	MITDB	ACC=98.4 SEN=91.1 SPF=98.7
Bouny et al.[16]	2020	ML-WCNN	MITDB	ACC=99.57
Sangaiah et al.[12]	2020	Wavelet coefficient features+HMM	MITDB	ACC=99.7 SEN=99.7
Jha et al.[13]	2020	Q-wavelet features+SVM	MITDB	ACC=99.27 SEN=96.22
Ge et al.[17]	2021	Squeeze-and-Excitation module	CPSCDB	ACC=83.0 SEN=82.7
Rahul et al.[14]	2021	Statistical features+SVM	MITDB	ACC=99.51 SEN=99.28
Zhang et al. [18]	2022	CNN+U-Net	CPSCDB	ACC=94.8 SEN=81.6
Nahak et al.[15]	2023	ECG features+SVM	MITDB	ACC=98.9
Zahid et al. [19]	2023	Self-operational network	MITDB	ACC=99.2 SEN=99.1
Park et al.[20]	2023	FCN+LSTM	MITDB	ACC=98.44
Cheng et al.[22]	2024	LFG-Net	CPSCDB	ACC=84.2
Sarankumar et al.[23]	2024	GRU+Autoencoder	CPSCDB	ACC=98.8 SEN=99.2
Yang et al.[24]	2025	U-Net+Transformer	CPSCDB	ACC=95.4
This research	2026	CNN+NDAMM	CPSCDB MITDB	CPSCDB ACC=99.4 SEN=99.0 SPF=99.5 MITDB ACC=99.2 SEN=98.8 SPF=99.4

two categories: machine learning-based methods [10, 15] and deep learning-based models [16, 24]. In terms of MLM, leveraging the conventional ML approaches to ECG signal identification is challenging because of the difficulties in inferior algorithm robustness, high-frequency presence of over-fitting and the deficiency of unified guidelines for feature extraction. Although research works Sangaiah et al.[12] and Rahul et al.[14] obtained 99.7% and 99.51% ACC, they still cannot avoid the above difficulties and have higher model complexity. These difficulties have become the main obstacles for conventional ML strategies to engage in practical applications. By comparisons, our strategy not only mitigates the problems but also reveals better performance.

Recent advancements in ECG pattern recognition have explored various deep learning methodologies with existing approaches employing diverse architectures including CNN blocks [16, 17], U-Net structures [18], and specialized frameworks such as self-operational networks [19], LFG-Net [22], auto-encoders [23] and transformer-based models [24]. Our results demonstrate that NDAMM achieves superior performance compared to these benchmark methods, with the exception of the CNN-based approach presented in [16] which attains a marginally higher accuracy of 99.57%. However, as detailed in Section 5.1, this performance advantage comes at the cost of significantly increased model complexity, particularly regarding parameter count and computational requirements. Through comprehensive evaluation metrics encompassing both classification accuracy and computational efficiency, our proposed NDAMM framework establishes an optimal balance between diagnostic performance and operational practicality. These comparative analyses confirm that NDAMM establishes an optimal balance between computational efficiency and diagnostic accuracy, achieving 99.4% classification accuracy while maintaining 48.1% fewer parameters and 60.0% lower FLOPs than conventional architectures.

5.4. Limitations

While this research demonstrates promising results, several limitations warrant discussion. First, clinical validation of NDAMM is currently confined to

PAC and PVC identification scenarios. Although this focused evaluation establishes foundational performance metrics, future investigations will extend the framework's applicability to broader arrhythmia classification tasks and diverse biological signal processing domains, including electromyography and electroencephalography analysis. Second, while the proposed architecture achieves favorable computational efficiency, opportunities exist for further optimization through alternative matrix decomposition strategies and adaptive parameter allocation mechanisms. Particularly, integrating hybrid attention modules with dynamic kernel selection could enhance feature representation capabilities without compromising computational economy. Third, the current implementation employs single-lead ECG analysis-extending this to multi-lead configurations and multimodal physiological data integration represents a critical direction for clinical translation. These planned extensions aim to validate NDAMM's generalizability across varied healthcare monitoring applications while maintaining its inherent computational advantages.

6. Conclusion

To compensate for the limitation of current DLM models ignoring information interaction, we construct NDAMM to distill interaction information from the aspects of both Spatial and Temporal dimensions. Multi-scale feature fusion strategy is utilized to adaptively select the optimal scope of neighbors participating in information interaction in spatial dimension and an atrous pyramidal pooling mechanism is adapted to generate the interaction probability for each agent in temporal dimension. The effectiveness of NDAMM is evaluated across PAC and PVC detection task. Experimental result achieves stable improvements over existing approaches and offers a good trade-off between performance and complexity. This study plans to extend NDAMM module's implementation in lightweight embedded systems, capitalizing on its exceptional computational efficiency and low memory footprint to enable high-performance edge computing applications.

7. Competing Interest

The authors declare no conflict of interest. This study is not a clinical trial and as Not applicable for registration details.

8. Publish declaration

Not applicable.

9. Ethics declaration

Not applicable.

10. Participate declaration

Not applicable.

11. Data and codes availability

All original ECG data are publicly available from both CPSCDB and MITDB at `https://physionet.org/content/challenge-2020/1.0.2/training/cpsc_2018/#files-panel` and `https://www.physionet.org/content/mitdb/1.0.0/`. These codes are available from the corresponding author upon reasonable request.

12. Acknowledgments

This work was funded by Anhui Provincial Department of Education Project (Grant No. 2025AHGXZK40256), Quality Engineering Project of Bengbu Medical University (Grant No. 2025jyxm39), '512' Outstanding Talents Fostering Project of Bengbu Medical College (Grant No. BY51201312) and Natural Science Youth Project of Bengbu Medical University (Grant No. 2024byzd133).

13. References

References

[1] Z.Y. Ge, H.Q. Cheng, Z. Tong, et al., A knowledge-driven graph convolutional network for abnormal electrocardiogram diagnosis, Knowl-Based Syst. 296 (2024) 111906.

[2] G.F. Ren, H. Liu, Y.W. Zhang, et al., Attack detection based on encoding-decoding approach for cyber-physical systems, IEEE Trans. Syst. Man Cy-S. 53 (10) (2023) 6293-6302.

[3] P. Singh, Y.P. Huang, AKDC: ambiguous kernel distance clustering algorithm for COVID-19 CT scans analysis, IEEE Trans. Syst. Man Cy-S. 54 (10) (2024) 6218-6229.

[4] K.C. Huang, W.J. Lyu, M.H. Yang, et al., PTT: point-trajectory transformer for efficient temporal 3D object detection, in: Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR), 2024.

[5] A. Anwar, Y. Khalifa, J.L. Coyle, et al., Transformers in biosignal analysis: a review, Inform. Fusion 114 (2025) 102697.

[6] H.H. Chen, X.X. Chu, Y.J. Ren, et al., PeLK: parameter-efficient large kernel convNets with peripheral convolution, in: Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR), 2024.

[7] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., An image is worth 16x16 words: Transformers for image recognition at scale, in: Proceedings of the IEEE International Conference on Learning Representation (ICLR), 2021.

[8] Z. Liu, Y.T. Lin, Y. Cao, et al., Swin transformer: hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2021.

[9] D. Wen, H.R. Xu, Z.C. He, et al., Density-adaptive model based on motif matrix for multi-agent trajectory prediction, in: Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR), 2024.

[10] W.L. Zhu, X.H. Chen, Y. Wang, et al., Arrhythmia recognition and classification using ECG morphology and segment feature analysis, IEEE Trans. Comput. Bi. 16 (1) (2019) 131-138.

[11] B.R. Oliveira, C.C.E. Abreu, M.A.Q. Duarte, et al., Geometrical features for premature ventricular contraction recognition with analytic hierarchy process based machine learning algorithms selection, Comput. Meth. Prog. Bio. 169 (2019) 59-69.

[12] 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.

[13] C.K. Jha, M.H. Kolekar, Cardiac arrhythmia classification using tunable Q-wavelet transform based features and support vector machine classifier, Biomed. Signal Proces. 59 (2020) 101875.

[14] 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.

[15] S. Nahak, A. Pathak, G. Saha, Fragment-level classification of ECG arrhythmia using wavelet scattering transform, Expert Syst. Appl. 224 (2023) 120019.

[16] L.E. Bouny, M. Khalil, A. Adib, An end-to-end multi-level wavelet convolutional neural networks for heart diseases diagnosis, Neurocomputing 417 (2020) 187-201.

[17] 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.

[18] X. Zhang, H. Wu, T. Chen, et al., Automatic diagnosis of arrhythmia with electrocardiogram using multiple instance learning: from rhythm annotation to heartbeat prediction, Artif. Intell. Med. 132 (2022) 102379.

[19] M.U. Zahid, S. Kiranyaz, M. Gabbouj, Global ECG classification by self-operational neural networks with feature injection, IEEE Trans. Biomed. Eng. 70 (1) (2023) 205-215.

[20] J. Park, K. Lee, N. Park, et al., Self-attention LSTM-FCN model for arrhythmia classification and uncertainty assessment, Artif. Intell. Med. 142 (2023) 102570.

[21] Y.Q. 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.

[22] Y.H. Cheng, D.Y. 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.

[23] R. Sarankumar, M. Ramkumar, K. Vijaipriya, Bidirectional gated recurrent unit with auto encoders for detecting arrhythmia using ECG data, Knowl-Based Syst. 294 (2024) 111696.

[24] H.X. Yang, Y.F. Zhao, X.W. Yang, PearsonUnet: ECG classification model integrating pearson related beats, IEEE Trans. Instrum. Meas. 74 (2025) 2508011.

[25] G.B. Moody, R.G. Mark, The impact of the MIT/BIH arrhythmia database, IEEE Eng. Med. Biol. Mag. 20 (3) (2001) 45-50.

[26] 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.

[27] J. Kim, J. Kong, J. Son, Conditional variational autoencoder with adversarial learning for end-to-end text-to-speech, in: Proceedings of International Conference on Machine Learning (ICML), 2021.

[28] R. Milo, S. Shen-Orr, S. Itzkovitz, et al., Network motifs: simple building blocks of complex networks, Science 298 (5594) (2022) 824-827.

[29] L.C. Chen, G. Papandreou, I. Kokkinos, et al., DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully Connected CRFs, IEEE Trans. Pattern Anal. 40 (4) (2018) 834-848.

[30] C.R. Qi, L. Yi, H. Su, et al., PointNet++: deep hierarchical feature learning on point sets in a metric space, in: Proceedings of Neural Information Processing Systems (NIPS), 2017.

[31] X.Q. Zhang, Y. Yuan, Q. Wang, ROI-wise reverse reweighting network for road marking detection, in: Proceedings of European Conference on Computer Vision (ECCV), 2018.

[32] K. He, X. Zhang, S. Ren, et al., Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE Trans. Pattern Anal. 37 (9) (2015) 1904-1916.

[33] Q. Tang, F.G. Liu, J. Jiang, et al., EPRNet: efficient pyramid representation network for real-time street scene segmentation, IEEE Trans. Intell. Transp. 23 (7) (2022) 7008-7016.

[34] R.Z. Cai, Z.T. Yu, C.Q. Kong, et al., S-Adapter: generalizing vision Transformer for face anti-spoofing with statistical tokens, IEEE Trans. Inf. Foren. Sec. 19 (2024) 8385-8397.

[35] P. Singh, A. Sharma, Interpretation and classification of arrhythmia using deep convolutional network, IEEE Trans. Instrum. Meas. 71 (2022) 2518512.

[36] S. Din, M. Qaraqe, O. Mourad, et al., ECG-based cardiac arrhythmias detection through ensemble learning and fusion of deep spatial-temporal and long-range dependency features, Artif. Intell. Med. 150 (2024) 102818.

[37] S. Yasuki, M. Taki, CAM back again: large kernel CNNs from a weakly supervised object localization perspective, in: Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR), 2024.