

# 크로스 어텐션 기반 ECG-RR 중간 융합을 통한 환자 간 부정맥 분류의 강건성 향상: 다중 백본 검증

\*김류하, \*허준영, \*차재빈, \*박영철†, \*\*조성필  
\*연세대학교 전산학과, \*\*\*(주)메쥬

## Improving Inter-Patient Arrhythmia Classification via Cross-Attention Based ECG-RR Intermediate Fusion: Multi-Backbone Validation

\*Ryuha Kim, \*Junyeong Heo, \*Jaebin Cha, \*Youngcheol Park†, \*\*Sungpil Cho  
\*Dept. of Computer Science, Yonsei University, \*\*Mezoo Co., Ltd.

### 요 약

웨어러블 기기의 확산으로 단일 리드 ECG 기반 실시간 부정맥 모니터링의 필요성이 증가하고 있으나, 환자 간(inter-patient) 평가 환경에서 딥러닝 모델의 일반화 성능 저하가 주요 제약으로 남아있다. 특히 심방성 조기박동(SVEB)은 정상 박동과 형태학적으로 유사하여 형태학 기반 분류에서 오분류가 빈번하다. 기존 연구들은 RR 간격 리듬 특징을 late fusion으로 결합하였으나, 표현 학습 과정에서 형태학-리듬 정보의 공동 최적화에는 한계가 있다. 본 연구는 멀티헤드 크로스 어텐션(MHCA)을 활용하여 RR 임베딩을 단일 토큰 쿼리로, ECG 특징 시퀀스를 Key/Value로 사용하는 중간 융합 방법을 제안한다. 제안 방법의 백본 독립적 일반화를 검증하기 위해 ResUNet과 MACNN-SE에서 동일한 전처리, 라벨 매핑, 평가 프로토콜로 실험을 수행하였다. MIT-BIH 부정맥 데이터베이스의 표준 AAMI DS1/DS2 환자 간 프로토콜에서, ResUNet 기반 B2 모델은 Macro F1 65.69%를, MACNN-SE 기반 하이브리드 모델은 Macro Sensitivity 70.83%를 달성하여, 두 백본 모두에서 baseline 대비 일관된 성능 향상을 확인하였다.

### Abstract

The proliferation of wearable devices has increased the demand for real-time arrhythmia monitoring using single-lead ECG; however, the degradation of deep learning model generalization in inter-patient evaluation remains a major constraint. In particular, supraventricular ectopic beats (SVEB) are frequently misclassified due to morphological similarity with normal beats. While prior studies have combined RR-interval rhythm features via late fusion, limitations remain in joint optimization of morphological and rhythm information during representation learning. This study proposes an intermediate fusion method using Multi-Head Cross-Attention (MHCA), where the RR embedding serves as a single-token query attending over the temporal ECG feature sequence (Key/Value). To validate backbone-independent generalization, experiments were conducted on ResUNet and MACNN-SE under identical preprocessing, label mapping, and evaluation protocols. Under the standard AAMI DS1/DS2 inter-patient protocol on the MIT-BIH Arrhythmia Database, the ResUNet-based B2 model achieved Macro F1 of 65.69%, and the MACNN-SE based hybrid model achieved Macro Sensitivity of 70.83%, confirming consistent improvements over baselines across both backbones.

**Keywords:** Electrocardiogram (ECG), Arrhythmia Classification, Cross-Attention, Intermediate Fusion

### I. 서론

웨어러블 기기의 확산으로 단일 리드 ECG 기반의 장시간 실시간 모니터링이 보편화되고 있다. 이에 따라 제한된 관측 채널에서도 부정맥을 자동으로 탐지하고 분류할 수 있는 기술의 필요성이 증가하고 있다[1]. 그러나 딥러닝 기반 모델은 학습에 포함되지 않은 새로운 환자에 대해 일반화해야 하는 환자 간(inter-patient) 환경에서 성능 저하가 반복적으로 보고되어 왔다[2].

특히 심방성 조기박동(SVEB)은 정상 박동과 QRS 형태가 유사하여 형태학적 기반 분류에서 오분류가 빈번하다[3]. 반면 SVEB는 조기 발생으로 인해 RR 간격이 단축되는 특징적인 리듬 변화를 동반하므로, RR 간격 정보는 형태학 기반 표현의 한계를 보완하는 중요한 단서가 된다. 기존 연구들은 RR 간격을 딥러닝 특징과 결합하여 SVEB 검출 성능 향상을

보고하였으나[4-6], 대부분 late fusion 또는 보조 입력 형태에 머물러 표현 학습 과정에서 형태학·리듬 정보를 충분히 공동 최적화하지 못하는 제약이 남는다.

본 연구에서는 멀티헤드 크로스 어텐션(MHCA)을 활용한 intermediate fusion 방법을 제안한다. RR 임베딩을 단일 토큰 쿼리로, ECG 백본에서 추출된 특징 시퀀스를 Key/Value로 사용하여, 리듬 정보가 형태학적 표현의 특정 시점을 동적으로 참조하도록 한다. 제안 방법의 백본 독립적 일반화 효과를 검증하기 위해, ResUNet과 MACNN-SE 두 가지 백본 구조에서 동일한 실험 조건으로 평가를 수행한다.

## II. 제안 방법

### 2.1 문제 정의

단일 리드 ECG 박동  $x$ 와 RR 간격 특징  $r$ 이 주어질 때, 부정맥 클래스  $y$ 를 예측하는 분류기  $f(x, r)$ 를 학습한다. 학습 목표는 다음 손실 함수의 최소화로 정의된다:

$$L = L_{CE} + \lambda \cdot L_{reg}$$

여기서  $L_{CE}$ 는 cross-entropy 손실,  $L_{reg}$ 는 weight decay 정규화 항이다.

### 2.2 크로스 어텐션 기반 중간 융합

ECG 백본 인코더는 입력  $x$ 로부터 형태학 특징 맵  $H$ 를 추출한다. RR 특징  $r$ 은 선형 임베딩을 통해  $d$  차원의 단일 토큰 쿼리  $Q$ 로 변환된다.

Attention shape 정의:

- Query:  $Q$  ( $B \times 1 \times d$ ) - RR 임베딩, 단일 토큰
- Key:  $K = H \cdot W_K$  ( $B \times T' \times d$ ) - ECG temporal sequence
- Value:  $V = H \cdot W_V$  ( $B \times T' \times d$ ) - ECG temporal sequence

크로스 어텐션은 다음과 같이 계산된다:

$$\text{Attn}(Q, K, V) = \text{softmax}(QK^T / \sqrt{d_k}) \cdot V$$

RR 임베딩이 단일 토큰 쿼리로 사용되어, ECG 특징 시퀀스의 temporal dimension  $T'$ 에 대해 attention을 수행한다.

### 2.3 RR 간격 기반 리듬 특징

표 1. RR 간격 기반 시간적 특징 정의

Feature	Formula	설명
Pre-RR	$RR_{i-1}$	직전 RR 간격
Post-RR	$RR_i$	직후 RR 간격
Local-RR	$\mu_l$	최근 10개 평균
Global-RR	$\mu_g$	전역 평균
Pre-Global diff	$RR_{i-1} - \mu_g$	편차
Post-Global ratio	$RR_{i-1} / \mu_g$	비율
Pre-Post ratio	$RR_{i-1} / RR_i$	Pre/Post 비율

## III. 실험 환경

### 3.1 데이터셋 및 평가 프로토콜

MIT-BIH 부정맥 데이터베이스[7]를 사용하였다. 모든 심박은 AAMI 권고[8]에 따라 5개 클래스(N, S, V, F, Q)로 재분류하였다. 환자 간 일반화 성능 평가를 위해 de Chazal et al.[4]이 제안한 표준 DS1/DS2 환자 단위 분할을 적용하였다. DS1(22명)은 학습용, DS2(22명)는 테스트용으로 사용된다.

표 2. 클래스별 샘플 분포

Class	DS1 (Train)	DS2 (Test)
N	45,842	44,238
S	944	1,837
V	3,788	3,221
F	415	388
Q	8	7

### 3.2 백본 구조 및 학습 설정

제안 융합 방법의 백본 독립적 일반화를 검증하기 위해 두 가지 백본에서 실험을 수행하였다:

- 1) ResUNet: ResU Block 기반 다중 해상도 특징 추출
- 2) MACNN-SE: Multi-scale Attention CNN with SE

모든 실험은 동일한 전처리, 라벨 매핑, optimizer, 평가 프로토콜을 공유하며, 백본 구조만 상이하다. 모든 결과는 5회 실험의 평균으로 보고한다.

## IV. 실험 결과 및 분석

### 4.1 ResUNet 백본에서의 융합 방법 비교

표 3. ResUNet 백본 기반 성능 비교 (%)

Method	Acc	Se	Pr	F1
Baseline	96.48	45.33	50.16	46.88
A1 (Concat)	95.70	47.58	38.02	41.60
A2 (Cross-Attn)	96.44	51.18	58.40	52.36
B1 (Concat-TF)	97.49	59.71	62.78	60.56
B2 (MHCA-TF)	97.72	62.90	73.62	65.69

Naive concatenation(A1)은 baseline보다 오히려 성능이 저하되었다. 이는 단순 연결이 feature scaling을 교란하고 adaptive alignment를 제공하지 못하기 때문으로 분석된다. 반면 Cross-Attention은 일관된 성능 향상을 보인다 (A2 > A1, B2 > B1).

### 4.2 MACNN-SE 백본에서의 융합 방법 비교

표 4. MACNN-SE 백본 기반 성능 비교 (%)

Method	Acc	Se	F1
Baseline	96.18	40.71	43.19
Naive Concat	93.22	47.56	43.84
MHCA (h=1)	92.48	43.28	45.41
Hybrid MHCA	93.74	70.83	58.54

### 4.3 Ablation Study: Hybrid MHCA 기여 분해

표 5. Hybrid MHCA Ablation (%)

Config	Early	MHCA	Se	$\Delta Se$
Baseline	O	X	40.71	-
MHCA only	X	O	43.28	+2.57
Hybrid	O	O	70.83	+30.12

MHCA 단독(+2.57%p)과 Early Fusion 단독의 합보다 Hybrid(+30.12%p)의 효과가 압도적으로 크다. 이는 Early Fusion이 제공하는 채널 수준 RR 정보와 Late MHCA가 제공하는 temporal attention이 시너지 효과를 발생시킴을 의미한다.

### 4.4 S 클래스(SVEB) 세부 분석

표 6. S 클래스(SVEB) 성능 비교 - ResUNet (%)

Method	S Pr	S Se	S F1
Baseline	21.22	11.21	14.67
B2 (MHCA)	70.45	66.58	68.46
$\Delta$	+49.23	+55.37	+53.79

크로스 어텐션 융합은 정상 박동과 형태학적으로 유사하여 오분류가 빈번한 SVEB에서 Precision과 Sensitivity를 동시에 대폭 향상시켰다.

## V. 결론

본 연구는 단일 리드 ECG 기반 부정맥 자동 분류에서 환자 간 일반화 성능을 향상시키기 위해, 멀티헤드 크로스 어텐션 기반 ECG-RR 중간 융합 방법을 제안하였다. RR 임베딩을 단일 토큰 쿼리로 사용하여 ECG temporal sequence에 대해 attention을 수행함으로써, 형태학-리듬 정보의 동적 상호작용을 가능하게 하였다.

ResUNet과 MACNN-SE 두 가지 백본에서 동일한 실험 조건으로 평가한 결과, 두 백본 모두에서 baseline 대비 일관된 성능 향상을 확인하였다. 특히 SVEB에서 크로스 어텐션 융합의 효과가 두드러졌다.

한계 및 향후 연구: 본 연구는 MIT-BIH 데이터베이스에서 검증되었으나, 대규모 실제 홀터 데이터셋에서의 추가 검증이 필요하다.

## REFERENCES

- [1] World Health Organization, "Cardiovascular diseases (CVDs)," WHO Fact Sheet, May 2023.
- [2] S. Mousavi and F. Afghah, "Inter- and intra-patient ECG heartbeat classification for arrhythmia detection," in Proc. IEEE ICASSP, pp. 1308-1312, May 2019.
- [3] H. Huang et al., "A new hierarchical method for inter-patient heartbeat classification using random projections and RR intervals," BioMed. Eng. OnLine, vol. 13, article no. 90, June 2014.
- [4] P. de Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," IEEE Trans. Biomed. Eng., vol. 51, no. 7, pp. 1196-1206, July 2004.
- [5] F. Zhou, Y. Sun, and Y. Wang, "Inter-patient ECG arrhythmia heartbeat classification network based on multiscale convolution and FCBA," Biomed. Signal Process. Control, vol. 86, article no. 105202, Sept. 2023.
- [6] G. Yan et al., "Fusing transformer model with temporal features for ECG heartbeat classification," in Proc. IEEE BIBM, pp. 898-905, Nov. 2019.
- [7] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH Arrhythmia Database," IEEE Eng. Med. Biol. Mag., vol. 20, no. 3, pp. 45-50, May-June 2001.
- [8] ANSI/AAMI EC57:1998/(R)2008, "Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms," 2008.