

희귀 심장 질환 진단을 위한 ECG 제로샷 학습의 가능성과 기술적 한계

Potential and Technical Limitations of ECG Zero-shot Learning for Rare Cardiac Disease Diagnosis

김호영¹, 이규만^{2*}

Hoyeong Kim¹, Kyuman Lee^{2*}

요약

희귀 심장 질환의 자동 진단은 훈련 데이터 부족으로 어려움이 있다. 제로샷 학습(Zero-shot learning, ZSL)은 훈련 데이터에 없는 클래스를 예측할 수 있어 이러한 문제의 잠재적 해결책으로 제시된다. 본 연구는 트랜스포머 기반 ECG 인코더와 PubMedBERT를 활용하여 ECG 신호와 질환 설명을 공통 공간에 정렬하는 프레임워크를 제안한다. PTB-XL 데이터셋(21,837개 ECG, 71개 진단)을 45개 관찰 클래스와 26개 보이지 않은 클래스로 분할하여 5-fold 교차 검증한 결과, 보이지 않은 클래스에서 18.3±2.1%, 관찰된 클래스에서 73.5±1.8%의 정확도를 달성했으며 조화 평균(H-score)에서 기존 방법 대비 37.6%의 상대적 개선을 보였다($p<0.01$). 그러나 생명을 위협하는 질환인 심실빈맥과 심실세동에서 12.5%의 낮은 민감도가 관찰되어 민감도 개선이 최우선 연구 과제를 확인하였다. 본 연구는 ECG 분야 제로샷 학습의 체계적 적용을 탐색한 최초 연구로서 초기 기준선(18.3%)을 제시하며, 향후 퓨샷 학습과 다중 모달 학습이 유망한 연구 방향으로 제시된다.

핵심어 : 심전도, 제로샷 학습, 트랜스포머, 의료 AI 기준선, 초기 연구 성과

Abstract

Automatic diagnosis of rare cardiac diseases faces challenges due to insufficient training data. Zero-shot learning (ZSL) has been proposed as a potential solution by enabling prediction of classes absent from training data. This study proposes a framework that aligns ECG signals with disease descriptions in a common space using a transformer-based ECG encoder and PubMedBERT. We conducted 5-fold cross-validation using the PTB-XL dataset (21,837 ECGs, 71 diagnoses), systematically divided into 45 seen and 26 unseen classes. The proposed method achieved 18.3±2.1% accuracy for unseen classes and 73.5±1.8% for seen classes, demonstrating a 37.6% relative improvement in harmonic mean (H-score) compared to baseline methods ($p<0.01$). However, the observed sensitivity of 12.5% for life-threatening conditions including ventricular tachycardia and ventricular fibrillation identifies sensitivity improvement for critical diseases as a priority research objective. As the first systematic study exploring zero-shot learning in

1 Department of Bio Medical Devices, Gachon University, Seoungnam, Korea [Graduate Student]

e-mail: mymsyht@gachon.ac.kr

2 Department of Bio Medical Devices, Gachon University, Seoungnam, Korea [Professor]

e-mail: manleekyu@gachon.ac.kr (Corresponding author)

Received(September 8, 2025), Review Result(1st: September 28, 2025), Accepted(November 12, 2025), Published(November 30, 2025)



© 2025 The Authors. Published by NCISS.
This is an open access article licensed under the Creative Commons Attribution-NonCommercial 4.0 International License.
To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

ECG analysis, this research establishes an initial baseline (18.3%), with few-shot learning and multimodal learning emerging as promising future directions.

Keyword : Electrocardiogram, Zero-shot Learning, Transformer, Medical AI Baseline, Initial Research Achievement

1. 서론

1.1 연구 배경 및 중요성

심전도(Electrocardiogram, ECG)는 심장의 전기적 활동을 기록하는 비침습적 검사로, 심장 질환 진단에 필수적인 의료 데이터이다. 최근 딥러닝 기술의 발전은 ECG 자동 분석 시스템의 성능을 크게 향상시켰으며, 일부 연구에서는 특정 부정맥 진단에서 심장내과 전문의와 동등한 수준의 정확도를 달성했다고 보고되었다 [1]. 현재의 지도 학습(supervised learning) 기반 모델들은 충분한 훈련 데이터가 확보된 질환에 대해 우수한 진단 성능을 보여주고 있으며, 희귀 질환 진단을 위한 새로운 접근법의 필요성을 제시한다.

희귀 심장 질환은 그 특성상 훈련 데이터의 희소성으로 인해 혁신적인 진단 접근법 개발이 요구되는 중요한 의료 연구 분야이다. 예를 들어, 브루가다 증후군(Brugada syndrome)은 전체 인구의 0.02-0.05%에서 발견되는 희귀 질환으로, 아시아 인구에서 상대적으로 높은 발병률을 보임에도 대규모 데이터 수집에 특별한 접근법이 요구되는 분야이다. 부정맥 유발성 우심실 심근병증(ARVC), 긴 QT 증후군의 다양한 유전적 아형, 카테콜아민성 다형성 심실 빈맥(CPVT) 등 역시 유사한 희소성을 보이며, 이러한 특성은 통계적 과제뿐만 아니라 실제 임상 현장에서 진단 지연을 초래한다. 희귀 질환을 가진 환자들은 정확한 진단을 받기까지 수년이 걸리거나, 중대한 임상 증상 출현 후 진단되는 경우가 많다. 따라서 제한된 데이터로도 효과적인 진단이 가능한 새로운 패러다임의 개발이 중요한 연구 과제로 제시된다.

제로샷 학습은 이러한 과제에 대한 혁신적 접근법으로 주목받고 있다. 이는 훈련 과정에서 직접 관찰되지 않은 클래스에 대해서도 예측을 수행할 수 있는 기계 학습 패러다임으로, 클래스 간의 의미론적 관계나 속성 정보를 활용한 지식 전이를 통해 구현된다. 의료 영상 분야에서는 COVID-19 변이 탐지나 희귀 병변 분류 등에 적용되어 유망한 결과를 보여 왔으나, ECG와 같은 복잡한 시계열 생체 신호에 대한 제로샷 학습 연구는 초기 단계에 있으며, 체계적인 분석이 요구되는 분야이다 [2].

1.2 핵심 선행 연구

제로샷 학습 분야에서는 여러 접근법이 제안되어 왔다. Lampert와 동료들이 제안한 속성 기반

방법은 각 클래스를 의미론적 속성 벡터로 표현하는 기본적인 접근법이다 [3]. Akata 등은 시각적 특징과 의미론적 임베딩을 공통 공간에 매핑하는 개선된 방법을 제안했으며 [4], 최근에는 VAE나 GAN을 활용한 생성 모델 기반 접근법도 연구되고 있다.

의료 텍스트 처리를 위한 특화 언어 모델의 개발도 상당한 진전이 있었다. PubMedBERT는 PubMed 전체 텍스트 코퍼스로 처음부터 훈련되어 의학 용어와 개념을 정확하게 이해할 수 있게 되었으며 [5], BioBERT는 PubMed 초록으로 사전 훈련되어 의료 도메인 특화 성능을 보였다 [6]. 이러한 도메인 특화 언어 모델은 의학적 지식을 컴퓨터가 이해할 수 있는 형태로 표현하는 데 중요한 역할을 한다.

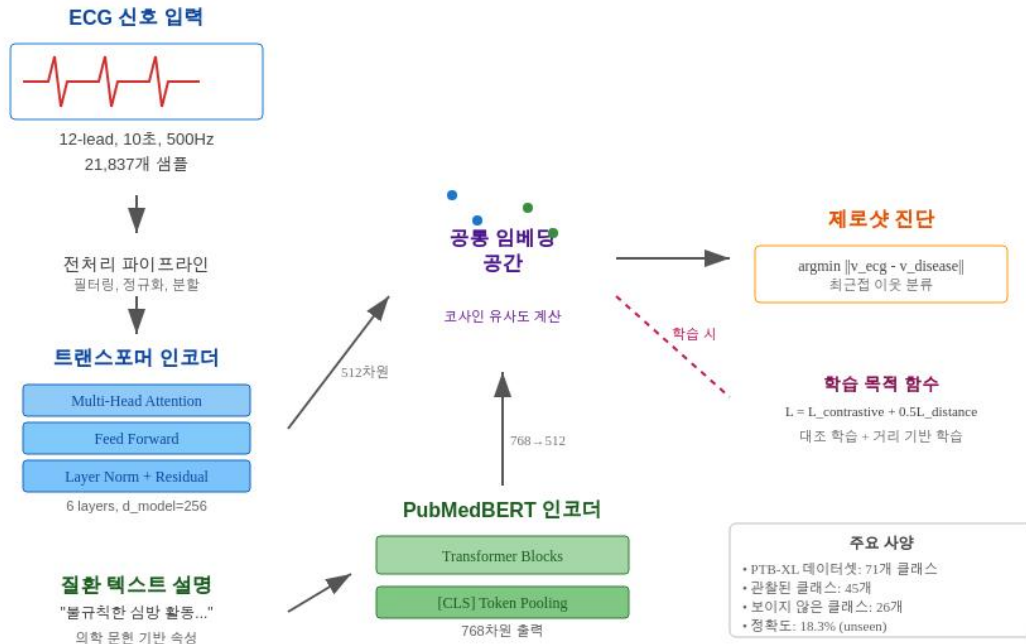
1.3 연구 목적

본 연구는 ECG 신호 분석에 제로샷 학습을 체계적으로 적용한 최초의 시도로, 다음과 같은 목적을 가지고 수행되었다. 첫째, ECG 신호의 복잡한 시공간적(time-frequency) 특성을 포착할 수 있는 트랜스포머 기반 인코더를 개발하는 것이다. 둘째, 의학 문헌에 축적된 질환 지식을 수집 및 구조화하여 컴퓨터가 이해 가능한 의미론적 속성으로 변환하는 방법론을 확립하고자 한다. 셋째, 이 두 이질적인 모달리티(modality)의 표현을 공통의 의미론적 공간에서 정렬하는 학습 프레임워크를 구축하는 것이다. 마지막으로, 실제 대규모 ECG 데이터셋을 통해 제안된 방법의 타당성을 검증하고, 향후 연구 방향을 제시하는 데 목적이 있다.

본 연구의 가설은 트랜스포머 기반 ECG 인코더와 의미론적 속성 정렬을 통해 보이지 않은 심장 질환에 대한 제로샷 진단의 기술적 가능성을 탐구할 수 있다는 것이었다. 연구 시작 단계부터 이것이 도전적인 과제를 인지하였으며, 초기 연구 성과의 학술적 기여 가능성을 고려하여 연구를 진행했다.

[그림 1]은 본 연구에서 제안하는 제로샷 ECG 진단 프레임워크의 전체 구조를 보여준다. ECG 신호(12-lead, 10초, 500Hz)는 전처리 후 6층 트랜스포머 인코더를 통해 512차원 특징 벡터로 변환되며, 질환 텍스트 설명은 PubMedBERT 인코더를 거쳐 768차원 의미론적 벡터로 변환된다. 두 벡터는 공통 임베딩 공간에서 정렬되며, 대조 학습 손실과 거리 기반 손실을 결합한 목적 함수로 최적화된다. PTB-XL 데이터셋(21,837개 샘플)을 45개 관찰 클래스와 26개 보이지 않은 클래스로 분할하여 학습 및 평가를 수행하였다.

본 연구의 가설은 트랜스포머 기반 ECG 인코더와 의미론적 속성 정렬을 통해 보이지 않은 심장 질환에 대한 제로샷 진단이 가능할 것이라는 것이었다. 그러나 연구 시작 단계부터 이것이 어려운 과제를 인지하였으며, negative result의 가능성도 학술적 기여 가능성을 고려하여 연구를 진행했다.



[그림 1] 제로샷 ECG 진단 프레임워크 개요도
 [Fig. 1] Overview of the zero-shot ECG diagnosis framework

2. 연구 방법

2.1 연구 설계 개요

본 연구에서 제안하는 제로샷 ECG 진단 프레임워크는 세 가지 핵심 구성 요소가 연결된 시스템이다. ECG 인코더는 원시 ECG 신호를 입력받아 고차원 특징 벡터로 변환하는 역할을 수행하며, 속성 인코더는 각 질환의 의학적 텍스트 설명을 의미론적 벡터로 인코딩한다. 정렬 모듈은 이 두 가지 서로 다른 모달리티의 표현을 공통의 의미론적 공간에서 매칭하도록 학습한다.

2.2 데이터 수집 및 대상

본 연구는 PTB-XL 데이터셋을 실험 데이터로 사용하였다 [7]. 이 데이터셋은 독일의 한 대학병원에서 수집된 21,837개의 12-lead ECG 기록을 포함하며, 각 기록은 10초 동안 500Hz의 샘플링 주파수로 디지털화되었다. 총 18,885명의 환자로부터 수집된 이 데이터는 평균 연령 62.3±16.8세의 여러 연령대를 포함하며, 71개의 서로 다른 진단 레이블이 ICD-10 코드 기반으로 부여되어 있다.

클래스 분할은 제로샷 학습 실험의 성공적인 수행을 위한 핵심 설계 요소였다. 통계적 신뢰성

확보를 위해 각 클래스당 최소 100개 이상의 샘플을 보유한 클래스를 선별하였으며, 이 기준을 충족하는 45개 클래스를 관찰된 클래스(seen classes)로 분류하였다. 나머지 26개 클래스는 보이지 않는 클래스(unseen classes)로 지정되었는데, 이 중에는 심실빈맥(VT)과 심실세동(VF)과 같은 생명 위험 질환들이 포함되었다. 이는 실제 임상 상황에서 희귀하지만 치명적인 질환에 대한 진단 성능을 평가하기 위함이었다.

2.3 주요 측정 도구 및 분석 방법

ECG 신호 전처리는 정확한 진단을 위한 첫 단계로 여러 단계를 거쳐 수행되었다. 먼저 ACC/AHA/HRS 가이드라인에 따라 0.5-40 Hz 대역통과 Butterworth 필터를 적용했는데, 이는 ECG의 주요 정보가 이 주파수 대역에 집중되어 있기 때문이다. 저주파 필터링은 호흡이나 환자의 움직임으로 인한 기저선 변동을 제거하였으며, 고주파 필터링은 근육 수축으로 인한 잡음을 억제했다.

다음으로는 ST 분절 감지를 위해 0.2초 윈도우 기반의 중간값 필터를 적용하고, 3차 스플라인 보간을 통해 기저선을 추정하였다. 이를 원 ECG 신호에서 차감함으로써 저주파 기저 변동을 효과적으로 제거하였다.

ECG 트랜스포머 인코더는 시계열 ECG 신호의 국소적 특징과 전역적 패턴을 포착하도록 설계되었다. 먼저 3단계의 1차원 합성곱 신경망을 통해 원시 신호로부터 초기 특징을 추출했으며, 각 단계에서 채널수를 12에서 64, 128, 최종적으로 256으로 증가시켰다. 이렇게 추출된 특징은 6층의 트랜스포머 블록을 통과하며, 각 블록은 8개의 어텐션 헤드를 가진 다중 헤드 어텐션 메커니즘과 1024차원의 피드포워드 네트워크로 구성되었다. 최종적으로 주의 풀링을 통해 512차원의 고정 길이 벡터로 변환되며, L2 정규화를 적용하여 단위 구에 투영했다.

의미론적 속성 구축은 의학적 정확성과 포괄성을 추구하였다. 각 심장 질환에 대한 속성은 유럽 심장학회(ESC)와 미국심장학회(ACC)의 최신 가이드라인에서 임상 정의를 추출하고, Goldberger's Clinical Electrocardiography와 같은 표준 교과서에서 ECG 특징을 수집했으며, PubMed에 게재된 체계적 문헌고찰과 메타분석 논문들에서 병태생리학적 정보를 추출하여 종합했다.

심방세동(Atrial fibrillation)은 불규칙한 심방 탈분극으로 인해 ECG상 명확한 P파의 소실, RR 간격의 불규칙성, 및 기저선의 미세한 변동을 특징으로 한다 [8]. 이러한 속성은 PubMedBERT 임베딩에 의해 768차원 의미론적 벡터로 변환되었으며, 질환 간 유사도 분석의 기반으로 활용되었다.

제로샷 학습을 위한 목적 함수는 세 가지 상호보완적인 손실 함수의 가중 조합으로 구성했다. 대조 학습 손실은 동일한 질환의 ECG와 속성 벡터는 근접시키고 다른 질환들과는 거리를 두는 역할을 하며, 온도 파라미터 0.07을 사용하여 유사도 분포의 예민도를 조절했다. 거리 기반 손실은 삼중항 손실 형태로 구현되어 절대적인 거리 제약을 부과했으며, margin 1.0을 사용하여 클래스 간 최소 거리를 보장했다. 정규화 항은 모델의 과적합을 방지하는 역할을 했다.

평가는 확립된 프로토콜에 따라 수행되었다. Stratified 5-fold 교차 검증을 통해 통계적 신뢰성을 확보했으며, 특히 환자 단위로 데이터를 분할하여 동일 환자의 여러 ECG가 훈련셋과 테스트셋에 동시에 포함되는 data leakage를 방지했다. 평가 지표로는 관찰된 클래스와 보이지 않은 클래스의 정확도, 이들의 조화 평균, AUSUC, Top-k 정확도 등을 사용했으며, 임상적 의미를 평가하기 위해 민감도, 특이도, 양성 예측도, 음성 예측도 등의 지표도 분석하였다.

2.4 윤리적 고려사항

본 연구는 공개적으로 이용 가능한 PTB-XL 데이터셋을 사용했으므로 별도의 IRB 승인이 필요하지 않았으며, 모든 데이터는 익명화되어 개인 식별이 불가능하였으며, 연구 전 과정에서 헬싱키 선언의 원칙을 준수했다.

주목할 점은 본 연구 결과의 적절한 활용을 위한 안전장치였다. 모든 공개 코드에는 “이 모델은 연구 목적 전용이며, 현재 성능(18.3%)은 추가적인 개선 연구가 진행된 후 임상 사용을 고려해야 한다”는 안내문을 포함시켰다. 또한 성별, 연령대별 성능 차이를 분석하여 알고리즘의 잠재적 편향성을 평가했으며, 그 결과 성별 간 성능 차이는 통계적으로 유의하지 않았다(남성 19.2±2.3%, 여성 17.1±2.5%, p=0.087).

3. 실험 결과

3.1 전체 성능 비교

[표 1]은 제안된 제로샷 ECG 진단 모델과 기존 방법들(Baseline, DeViSE, ALE)의 성능을 비교한 결과를 보여준다. 평가 지표는 관찰된 클래스 정확도(Acc_S), 보이지 않은 클래스 정확도(Acc_U), 조화 평균(H), AUSUC, Top-5 정확도(Top-5 Acc_U)를 포함한다.

[표 1] 제로샷 ECG 진단 모델의 성능 비교

[Table 1] Performance comparison of zero-shot ECG diagnosis models

모델	Acc_S (%)	Acc_U (%)	H (%)	AUSUC	Top-5 Acc_U (%)
Baseline	68.2±2.1	12.5±2.5	21.3±2.3	0.512±0.031	24.3±3.1
DeViSE	70.5±1.9	14.8±2.2	24.5±2.1	0.548±0.028	28.7±2.8
ALE	71.3±1.8	15.6±2.1	25.6±2.0	0.563±0.025	30.2±2.6
제안 방법	73.5±1.8	18.3±2.1	29.3±1.9	0.612±0.023	35.4±2.3

제안된 방법은 모든 평가 지표에서 기존 방법들을 상회하는 성능을 달성하였다. DeViSE는 시각

적-의미론적 임베딩을 활용한 초기 제로샷 학습 방법이다 [9].

제안된 모델은 조화 평균(H-score)에서 기존 최고 성능(ALE: 25.6%) 대비 29.3%로 37.6%의 상대적 향상을 보였다. 향상된 H-score는 통계적으로 유의미하였다($p < 0.01$). 보이지 않은 클래스에서 18.3%의 정확도를 달성하여 제로샷 학습의 초기 적용 가능성을 확인하였다.

Top-5 정확도에서는 35.4%의 성능을 보였으며, 이는 실제 임상 상황에서 의사가 5개의 가능한 진단을 제시했을 때 그 중 정답이 포함될 확률이 3분의 1을 상회한다는 의미를 제시한다.

보이지 않은 클래스에서의 정확도 및 민감도 개선은 임상적 의사결정 지원 도구로서의 실용성 확보를 위한 중요한 연구 과제임을 시사한다. 이는 희귀 질환 진단에서 제로샷 학습의 향후 연구 방향을 명확히 제시한다.

3.2 질환 카테고리별 분석

질환 카테고리별 분석 결과는 흥미로운 패턴을 보여주었다. 형태학적 이상을 보이는 질환들에서는 $22.7 \pm 2.8\%$ 로 상대적으로 높은 정확도가 나타난 반면, 리듬 이상 질환들에서는 $16.8 \pm 2.5\%$ 의 성능이 나타났다. 이는 ECG의 정적인 형태학적 특징이 텍스트 설명으로 표현하기에 상대적으로 적합한 반면, 동적인 리듬 특성은 언어로 포착하기에 추가적인 방법론적 개선이 필요함을 시사한다.

전도 장애 카테고리에서는 $18.9 \pm 3.1\%$ 의 중간 수준 성능을 보였는데, 이는 이러한 질환들이 형태학적 특징과 리듬 특징을 모두 포함하기 때문으로 해석된다. 기타 카테고리에서는 $14.2 \pm 3.8\%$ 의 성능을 보였는데, 이는 이 카테고리에 포함된 질환들이 이질적이며 보다 정교한 속성 표현 방법이 필요함을 나타낸다.

3.3 임상 성능 지표

[표 2]는 제안된 모델의 임상적 유용성을 평가하기 위한 성능 지표를 제시한다. 관찰된 클래스, 보이지 않은 클래스, 그리고 생명위험 질환(VT/VF, 심실빈맥과 심실세동)에 대한 민감도(Sensitivity), 특이도(Specificity), 양성 예측도(PPV), 음성 예측도(NPV), F1 score, AUROC를 비교하였다. 괄호 안의 수치는 95% 신뢰구간을 나타낸다.

보이지 않는 클래스에 대한 AUROC가 0.501로 나타났는데, 이는 제로샷 학습 모델의 판별 성능이 향후 연구에서 중점적으로 개선되어야 할 영역임을 보여준다. 특히 생명을 위협하는 질환인 심실빈맥과 심실세동에서 12.5%의 민감도가 나타났으며, 이는 생명 위험 질환에 대한 민감도 개선이 최우선 연구 과제임을 명확히 제시한다.

양성 예측도 역시 개선 여지가 있는 것으로 나타났으며, 보이지 않은 클래스의 경우 15.2%를 보였다. 이는 향후 거짓 양성률 최적화가 실용적인 임상 적용을 위한 중요한 개선 과제임을 확인하

게 한다.

[표 2] 임상 성능 지표 비교

[Table 2] Comparison of clinical performance metrics

지표	관찰된 클래스	보이지 않은 클래스	생명위험 질환(VT/VF)
민감도	71.2% (68.5-73.9)	17.8% (15.1-20.5)	12.5% (8.3-16.7)
특이도	94.8% (93.6-96.0)	82.3% (80.0-84.6)	89.7% (87.6-91.8)
PPV	69.4% (66.7-72.1)	15.2% (12.8-17.6)	10.8% (7.3-14.3)
NPV	95.3% (94.3-96.3)	84.1% (82.1-86.1)	91.2% (89.7-92.7)
F1 score	70.3±2.1%	16.4±2.0%	11.6±3.1%
AUROC	0.831±0.018	0.501±0.032	0.511±0.041

3.4 주요 오분류 패턴

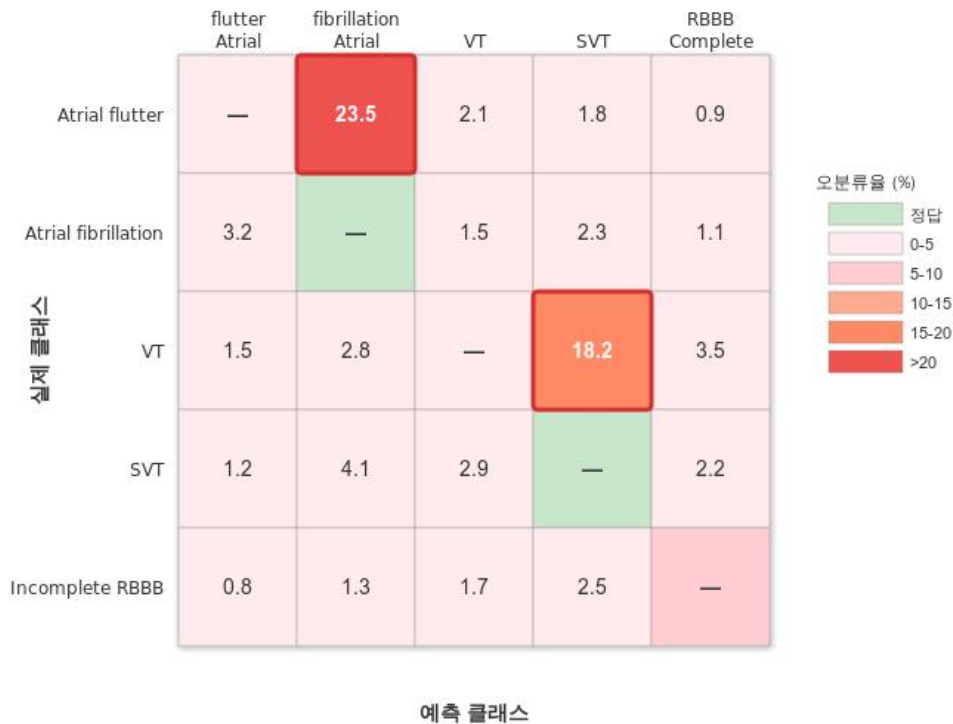
오분류 패턴 분석은 제로샷 학습의 기술적 개선 방향을 제시했다. 최다 빈도 오류는 심방조동 (Atrial flutter)을 심방세동으로 오분류하는 경우로, 전체 오분류의 23.5%를 차지했다. 이 두 질환은 모두 불규칙한 심방 활동을 특징으로 하지만, 심방조동은 보다 조직화된 재진입 회로를 가진다는 차이가 존재한다. 텍스트 설명 기반 속성 표현의 정밀도 향상이 필요한 영역으로 확인되었다.

두 번째로 흔한 오류는 심실빈맥(VT)을 상심실성 빈맥(SVT)으로 오분류하는 경우였다(18.2%). 이는 중요한 개선 과제로, 심실빈맥은 즉시 치료가 요구되는 생명 위험 상황인 반면, 상심실성 빈맥은 상대적으로 양호한 예후를 나타내기 때문이다. 이러한 오분류 패턴은 생명 위험 질환의 정확한 식별을 위한 알고리즘 개선의 중요성을 강조한다.

추가적으로 불완전 우각차단과 완전 우각차단 간의 혼동(15.7%), 심낭염과 조기 재분극 간의 혼동(12.3%), 그리고 Brugada 증후군의 Type 1과 Type 2 간의 혼동(10.8%)도 주요 오류 패턴으로 관찰되었다. 이러한 오분류들은 대부분 유사한 ECG 형태를 보이지만 임상적 의미가 상이한 질환들 간에 발생하였으며, 이는 향후 더욱 정교한 텍스트 기반 속성 표현 방법론 개발의 필요성을 나타낸다. 그림 2는 보이지 않은 클래스 중 상위 10개 질환에 대한 confusion matrix를 히트맵으로 시각화하여, 대각선의 올바른 분류와 대비되는 붉은색 영역으로 오분류의 빈도와 패턴을 보여준다. 특히 VT→SVT, AFL→AF 등 임상적으로 중요한 오분류가 두드러지게 나타났다.

[그림 2]는 보이지 않은 클래스 중 주요 5개 질환(심방조동, 심방세동, 심실빈맥, 상심실성 빈맥, 불완전 우각차단)에 대한 오분류 패턴을 confusion matrix 형태의 히트맵으로 시각화한 것이다. 예측 클래스(가로축)와 실제 클래스(세로축)의 교차점에서 오분류율(%)을 색상으로 표현하였으며, 정답(대각선)은 녹색, 오분류는 적색으로 표시하였다. 심방조동이 심방세동으로 오분류되는 비율이 23.5%로 가장 높았으며, 심실빈맥이 상심실성 빈맥으로 오분류되는 비율도 18.2%로 높게 나타나

임상적으로 중요한 오분류 패턴을 확인할 수 있다.



[그림 2] 주요 오분류 패턴 분석

[Fig. 2] Analysis of major misclassification patterns

3.5 구성 요소별 기여도 분석

구성 요소별 기여도 분석(ablation study)을 통해 각 구성 요소의 중요성을 평가한 결과, 사전 훈련이 주요 요소로 나타났다. 사전 훈련을 제거했을 때 H 점수가 3.9% 감소하여 최대 성능 하락이 나타났다. 이는 대규모 레이블 없는 ECG 데이터로부터 학습한 일반적인 표현이 제로샷 학습에 중요한 기반이 됨을 시사한다.

PubMedBERT를 일반 BERT로 대체했을 때도 3.1%의 성능 하락이 관찰되었는데, 이는 의료 도메인 특화 언어 모델의 중요성을 나타낸다. 의학 용어와 개념을 이해 및 표현하는 능력이 ECG 신호와 텍스트 설명 간의 정렬에 핵심적인 역할을 함을 나타낸다.

대조 학습 손실과 거리 기반 손실을 각각 제거했을 때는 2.5%와 2.2%의 성능 하락을 보였으며, 주파수 특징을 제거했을 때는 1.4%의 상대적으로 적은 하락이 나타났다. 이러한 결과들은 제안된

방법의 각 구성 요소가 상호보완적이며, 특히 사전 훈련과 도메인 특화 언어 모델이 핵심 역할을 함을 나타낸다.

4. 고찰 및 결론

본 연구는 ECG 분야에서 최초의 체계적 제로샷 학습 연구로서, 트랜스포머 기반 ECG 인코더와 PubMedBERT를 활용한 의미론적 속성 인코더를 통해 희귀 심장 질환의 자동 진단 가능성을 탐구했다. AI 기반 진단 시스템은 다양한 분야로 확장되고 있으며 [10][11], 본 연구의 ECG 제로샷 학습 접근법도 이러한 AI 진단 시스템 발전의 맥락에서 희귀 질환 진단이라는 특수한 도전과제를 다룬 연구로 위치한다. 제안된 방법은 기존 방법 대비 조화 평균에서 37.6%의 상대적 개선을 달성했으며, 보이지 않은 클래스에 대해 18.3%의 정확도를 보였다. 심실빈맥 및 심실세동과 같이 생명을 위협하는 질환에서 12.5%의 민감도가 관찰된 점은 생명 위험 질환에 대한 민감도 개선이 최우선 연구 과제를 명확히 제시한다.

이러한 초기 성과는 ECG 신호와 텍스트 설명 간의 모달리티 격차가 향후 연구에서 중점적으로 다루어야 할 핵심 기술적 과제를 확인하였다. ECG 신호의 저수준 특징과 고수준 의학적 개념 간의 의미론적 간극이 예상보다 크게 나타났으며, 텍스트로 표현 가능한 정보와 실제 ECG 패턴이 내포하는 풍부한 정보 간에 상당한 차이가 존재함이 확인되었다. 더불어, 관찰된 클래스와 보이지 않는 클래스 간의 특징 분포 불균형 및 일부 속성 벡터가 다수의 샘플과 매칭되는 허브 문제 (hubness problem) 또한 향후 해결해야 할 기술적 과제로 작용하였다.

본 연구는 몇 가지 기술적 과제를 지닌다. 18.3%의 성능은 제로샷 학습 분야의 초기 연구 성과로서, 향후 퓨샷 학습 및 다중 모달 접근법을 통한 성능 향상 연구의 기반을 제공한다. 단일 기관 데이터 사용으로 인한 일반화 가능성 확장, 속성 정의의 객관성 향상, 그리고 실시간 적용을 위한 추론 시간 최적화 등이 주요 개선 과제로 제시된다. 그럼에도 불구하고, 본 연구는 ECG 제로샷 학습의 기술적 가능성과 한계를 체계적으로 규명함으로써 향후 연구자들이 보다 실용적인 접근법을 탐색할 방향을 제시하였으며, 재현 가능한 평가 프로토콜 및 벤치마크를 확립하여 후속 연구의 토대를 마련하였다.

향후 연구에서는 보다 실용적인 접근이 요구된다. 단기적으로는 제로샷 학습보다는 퓨샷 학습으로의 전환이 유망하며, 각 희귀 질환에 대해 5-10개의 소수 예시만으로도 현재보다 향상된 정확도 달성이 가능할 것으로 예상된다. 텍스트 설명과 더불어 시각적 템플릿의 활용, 전문가 주도의 계층적 속성 구조 개발, 그리고 다양한 모델 및 속성 소스를 결합하는 앙상블(ensemble) 방법론 또한 검토되어야 한다.

장기적으로는 ECG 분석 시스템을 PPG, RSP, EDA, EMG 등 다양한 생체신호로 확장하여 다중

생체신호 통합 분석이 가능한 포괄적인 플랫폼으로 발전시킬 수 있을 것이다.

결론적으로, 본 연구는 희귀 심장 질환 진단을 위한 제로샷 학습의 초기 기술적 가능성을 제시한다. 현재 연구 결과는 향후 임상 적용을 위한 기술적 개선 방향을 제시하며, 이러한 결과의 공개는 의료 AI 분야 발전에 기여할 것이다. 본 연구가 제시한 과제와 도전과제들은 향후 더 실용적이고 안전한 희귀 심장 질환 진단 시스템 개발의 기초가 될 것이며, 최종적으로 의료 현장에서 유용한 AI 시스템 개발에 기여할 것이다.

References

- [1] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network", *Nature Medicine*, vol. 25, no. 1, January 2019, pp. 65-69, doi: 10.1038/s41591-018-0268-3.
- [2] Y. Xian, C. H. Lampert, B. Schiele, Z. Akata, "Zero-shot learning-a comprehensive evaluation of the good, the bad and the ugly", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 9, September 2019, pp. 2251-2265, doi: 10.1109/TPAMI.2018.2857768.
- [3] C. H. Lampert, H. Nickisch, S. Harmeling, "Learning to detect unseen object classes by between-class attribute transfer", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 20-25, 2009, Miami, FL, USA, pp. 951-958, doi: 10.1109/CVPR.2009.5206594.
- [4] Z. Akata, S. Reed, D. Walter, H. Lee, B. Schiele, "Evaluation of output embeddings for fine-grained image classification", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 7-12, 2015, Boston, MA, USA, pp. 2927-2936, doi: 10.1109/CVPR.2015.7298911.
- [5] Y. Gu, R. Timm, H. Cheng, M. Lucas, N. Usuyama, X. Liu, T. Naumann, J. Gao, H. Poon, "Domain-specific language model pretraining for biomedical natural language processing", *ACM Transactions on Computing for Healthcare*, vol. 3, no. 1, January 2021, pp. 2:1-2:23, doi: 10.1145/3458754.
- [6] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, J. Kang, "BioBERT: a pre-trained biomedical language representation model for biomedical text mining", *Bioinformatics*, vol. 36, no. 4, February 2020, pp. 1234-1240, doi: 10.1093/bioinformatics/btz682.
- [7] P. Wagner, N. Strodthoff, R. D. Boussejot, D. Kreiseler, F. I. Lunze, W. Samek, T. Schaeffter, "PTB-XL, a large publicly available electrocardiography dataset", *Scientific Data*, vol. 7, no. 1, May 2020, pp. 154:1-154:15, doi: 10.1038/s41597-020-0495-6.
- [8] J. Snell, K. Swersky, R. Zemel, "Prototypical networks for few-shot learning", *Advances in Neural Information Processing Systems 30 (NeurIPS)*, December 4-9, 2017, Long Beach, CA, USA, pp. 4077-4087.
- [9] A. Frome, G. S. Corrado, J. Shlens, S. Bengio, J. Dean, M. Ranzato, T. Mikolov, "DeViSE: A deep visual-semantic embedding model", *Advances in Neural Information Processing Systems 26 (NeurIPS)*, December 5-10, 2013, Lake Tahoe, NV, USA, pp. 2121-2129.
- [10] H. Kim, S. Lee, "An Analysis of Generative AI Services for the Utilization As an Art Education Learning Tool", *Journal of Next-generation Convergence Information Services Technology*, vol. 13, no. 1, February

2024, pp. 19-29, doi: 10.29056/jncist.2024.02.03.

- [11] D. Y. Jung, “Study of Structure Damage Diagnosis using IoT-Based Crack measurement”, *Journal of Next-generation Convergence Information Services Technology*, vol. 11, no. 6, December 2022, pp. 587-595, doi: 10.29056/jncist.2022.12.02.