



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사 학위논문

딥러닝을 활용한 종단 심전도 분석:
좌심실 박출률 감소 감지 성능 개선

Longitudinal ECG Analysis
with Deep Learning for Improved Detection of
Reduced Left Ventricular Ejection Fraction

2025 년 2 월

서울대학교 대학원

협동과정 바이오엔지니어링 전공

정 수 민

딥러닝을 활용한 종단 심전도 분석: 좌심실 박출률 감소 감지 성능 개선

지도 교수 윤 형 진

이 논문을 공학석사 학위논문으로 제출함
2025 년 02 월

서울대학교 대학원
협동과정 바이오엔지니어링
정 수 민

정수민의 공학석사 학위논문을 인준함
2024 년 12 월

위 원 장 최 진 욱 (인)

부위원장 윤 형 진 (인)

위 원 김 광 수 (인)

초 록

연구 배경 : 심부전은 전 세계적으로 심각한 건강 문제이며, 좌심실 수축 기능 저하는 그 주요 원인 중 하나이다. 기존의 심초음파 검사는 검사자 의존적이고 주관적인 해석의 한계가 있으며, 단일 시점 심전도 기반 진단 연구들은 시간에 따른 변화와 개인별 특성을 고려하지 못하는 제한점이 있었다.

연구 방법 : 본 연구에서는 환자의 종단 심전도 데이터를 활용하여 좌심실 박출률 감소를 예측하는 딥러닝 모델을 개발하였다. 제안된 가중합 기반 프레임워크는 다중 방문 환자의 원본 심전도 신호를 입력으로 사용하며, 다양한 모델 아키텍처에 적용 가능하도록 설계되었다. 지수 가중 함수를 통해 시점 간 간격에 따른 가중치를 할당하여 최신 데이터의 중요성을 강조하면서도 과거 데이터의 정보를 효과적으로 활용하였다.

연구 결과 : 제안된 종단 방법이 단일 시점 방법보다 일관되게 우수한 성능을 보였으며, 특히 ECGNet 모델이 가장 높은 성능을 달성하였다. 심전도 검사 횟수 증가에 따른 성능 향상을 통해 종단 데이터 활용의 효과성을 입증하였다. 기계학습 모델과의 비교 분석에서도 제안된 딥러닝 접근법의 우수성을 확인하였다. Expected Gradients 기법을 활용한 모델 해석 가능성 분석을 통해 모델의 의사결정 과정에 대한 임상적 통찰을 제공하였다.

결론 : 본 연구는 딥러닝과 종단 심전도 데이터 분석을 통해 좌심실 박출률 감소를 예측하는 모델을 개발하였다. 제안된 모델은 심부전 환자의 조기 진단, 예후 예측, 맞춤형 치료 계획 수립 등에 활용될 수 있을 것으로 기대된다. 그러나 단일 기관 데이터 사용과 클래스 불균형

등의 한계점이 존재한다. 향후 다기관 연구, 다중 모달리티 데이터 통합, 다양한 심혈관 질환으로의 적용 확장 등을 통해 모델의 일반화 가능성과 임상적 유용성을 더욱 높일 수 있을 것으로 기대된다.

주요어 : 심부전, 좌심실 박출률 감소, 심전도, 딥러닝, 종단 데이터 분석

학 번 : 2022-27947

목 차

제 1 장 서론	1
1.1 연구의 배경 및 필요성	1
1.2 문제 제기 및 연구 목적	2
제 2 장 연구방법	5
2.1 연구 대상자	5
2.2 데이터 수집	9
2.3 프레임워크 개발	11
2.4 모델 학습 및 평가과정	19
2.5 통계 분석	22
제 3 장 연구결과	23
3.1 데이터 구성	23
3.2 Cross-sectional vs. Longitudinal 성능 비교	25
3.3 기계학습 모델을 활용한 성능 비교	27
3.4 예측 결과에 대한 해석 가능성	31
3.5 심전도 개수에 따른 성능 변화 분석	33
3.6 파라미터에 따른 성능 분석	35
3.7 클래스 불균형: 데이터 처리 기법의 영향 분석	38
3.8 추가 분석	42
제 4 장 고찰	46
4.1 실험 결과 고찰	46
4.2 한계점 및 발전 방향	48
제 5 장 결론	50
참고문헌	52
Abstract	58

표 목차

Table 1. Hyperparameter tuning set	21
Table 2. Baseline characteristics	24
Table 3. Performance comparison between longitudinal and cross-sectional setting	26
Table 4. Performance comparison between longitudinal and cross-sectional settings on machine learning models.....	29
Table 5. Performance comparison between the best performing deep learning and machine learning models in longitudinal settings	30
Table 6. Performance per hyperparameter value.....	37
Table 7. Sensitivity analysis according to different case:control ratios	40
Table 8. Sensitivity analysis according to different augmentation ratios	41
Table 9. Performance comparison between Constant and Change groups in LVEF status	44

그림 목차

Figure 1. Cohort flow chart	6
Figure 2. Data selection process example	8
Figure 3. Example of ECHO report and ECG signal	10
Figure 4. Overview of the Longitudinal ECG framework.....	12
Figure 5. Weighted summation locations in ECGNet	14
Figure 6. Example of equally sampling strategy.....	15
Figure 7. Examples of exponential weighting functions.....	17
Figure 8. Model explainability	32
Figure 9. The impact of the number of ECGs on model performance.....	34
Figure 10. Comparison of embedding mean values based on ECG record count.....	45

List of Abbreviations

LVSD: Left Ventricular Systolic Dysfunction

LVEF: Left Ventricular Ejection Fraction

ECHO: Echocardiogram

ECG: Electrocardiogram

AI: Artificial Intelligence

XML: eXtensible Markup Language

Hz: Hertz

RBBB: Right bundle branch

LBBB: Left bundle branch

RMS: Root mean square

LF/HF ratio: Low Frequency to High Frequency ratio

CNN: Convolution Neural Network

ViT: Vision transformer

AUROC: Area Under the Receiver Operating Characteristic
curve

AUPRC: Area Under the Precision-Recall Curve

PPV: Positive Predictive Value

NPV: Negative Predictive Value

TP: True Positive

FP: False Positive

TN: True Negative

FN: False Negative

BMI: Body Mass Index

LDA: Linear Discriminant Analysis

제 1 장 서론

1.1 연구의 배경 및 필요성

심부전은 전 세계적으로 심각한 건강 문제로 대두되고 있다. 전 세계적으로는 최소 6,400 만 명 이상이 심부전을 앓고 있는 것으로 추정되며, 미국 인구의 약 2.3%가 심부전으로 고통받고 있다 [1, 2]. 좌심실 수축 기능 저하(Left Ventricular Systolic Dysfunction; LVSD)는 심부전의 주요 원인 중 하나로 [3], 전신 색전증, 뇌졸중, 사망 위험을 증가시킨다 [4]. 따라서 좌심실 수축 기능 저하의 조기 진단과 적절한 관리는 심부전 환자의 예후 개선에 매우 중요하다 [5]. 특히, 좌심실 박출률(Left Ventricular Ejection Fraction; LVEF) 감소 여부를 확인하는 것은 LVSD 진단에서 중요하다. 좌심실 박출률은 심실이 수축할 때 펌프질되는 혈액의 비율을 나타내며, 수식 1 의 공식을 통해 계산된다. 이는 심실 기능의 척도로 심장의 펌프 기능을 평가하는 중요한 지표이며 질병의 진행 및 치료 반응을 정량화하고[6-8], 사망률을 독립적으로 예측하는 데에도 사용된다[9]. 여러 선행 연구에 따르면 좌심실 박출률 감소에 의해 심부전 및 관련 합병증의 위험이 증가하는 것으로 나타났다[10-13].

$$\text{좌심실 박출률(LVEF)} = \frac{\text{좌심실에서 펌프질된 혈액의 양}}{\text{좌심실에 있는 총 혈액의 양}} \times 100 \quad (1)$$

심초음파 검사(Echocardiogram; ECHO)는 현재 좌심실 박출률 이상 여부를 진단하는 표준 도구로 사용되고 있다. 그러나 심초음파검사사는 검사자 의존적이고 해석이 주관적이라는 제한점이 있다. 검사 결과의 정확성은 검사자의 전문성에 크게 좌우되며, 이는 자원이 제한된 의료

환경에서 심초음파 검사의 일상적인 사용을 제한한다 [7, 14]. 게다가 심초음파 검사가 지연되는 경우가 있어, 심부전 환자의 적절한 관리가 지연될 수 있다 [15].

이에 따라 심전도(Electrocardiogram; ECG) 등을 활용한 대안적 선별 도구 개발이 시도되어 왔다 [16-19]. 심전도는 심장의 전기적 활동을 기록하여 심장의 상태를 평가할 수 있는 도구로, 환자에게 불편을 주지 않으면서 빠르고 쉽게 실시할 수 있다는 장점이 있다. 심전도 기반 좌심실 수축 기능 저하 진단 연구가 1996 년부터 지속적으로 이루어져 왔다 [19]. 초기에는 심전도상의 단순한 이상 소견을 식별하는 수준이었으나, 최근에는 다양한 AI (artificial intelligence) 알고리즘이 개발되어 다양한 좌심실 수축 기능 저하 기준 (e.g., LVEF <35% [20-22], <40% [18, 23-25], or <50% [18, 26, 27])에 따른 좌심실 수축 기능 저하 진단에 활용되고 있다.

1.2 문제 제기 및 연구 목적

심부전은 만성적이고 진행성 질환으로, 심부전 환자의 경우 장기적인 심전도 모니터링이 권장된다. 반복적인 심전도 측정은 심부전의 진행 양상을 파악하고 치료에 대한 반응을 모니터링하는 데 유용하며, 이는 진단 및 예후 평가에 중요한 역할을 한다 [28]. 또한 실제 의료 현장에서도 심부전 환자의 지속적인 병원 방문과 반복적인 심전도 측정은 일반적이다. 그러나 심부전을 예측하는 기존 AI 알고리즘들은 훈련 샘플의 독립성을 가정하므로, 반복 측정된 데이터를 활용하지 않는다 [29].

따라서 심부전 환자의 장기적인 심전도 모니터링 데이터를 활용할 수 있는 AI 알고리즘 개발이 필요하다. 종단 심전도 분석 모델은 심부전

환자의 장기적인 심전도 데이터를 활용하여, 시간에 따른 심장 기능의 변화를 예측할 수 있다. 이는 심부전의 조기 진단 및 예후 예측, 치료 계획 수립 등에 활용될 수 있다. 또한, 종단 심전도 분석 모델은 환자 개인의 특성을 고려할 수 있으므로, 보다 정확하고 개인화된 예측이 가능하다. 따라서 심부전 환자의 관리에 있어 종단 심전도 분석 모델의 개발과 활용이 필수적이라 할 수 있다.

종단 심전도 데이터를 활용한 두 가지 선행 연구가 존재한다. [29]에서는 종단 심전도 데이터와 선형 혼합 모델(Linear Mixed Model)을 활용하여 혈중 칼륨 농도를 예측하는 모델을 개발하였다. 연구의 목표는 환자의 제일 최근 시점의 혈중 칼륨 농도를 예측하는 것이며, 이를 위해 이전 시점에서의 심전도-칼륨 농도 값 쌍들과 선형 혼합 모델을 활용하여 최종 예측 값을 보정하였다. 그러나 이러한 접근 방식에는 한계점이 존재한다. 환자의 맨 마지막 시점의 칼륨 농도를 예측하고자 하더라도 이전 시점의 모든 심전도 데이터에 대해서 쌍을 이루는 칼륨 농도 값이 존재해야만 정확한 예측이 가능하여 코호트가 제한적이다. 또 다른 선행 연구에서는 급성 심부전 예측을 위해 종단 심전도 데이터를 활용하였으나 [30] 몇 가지 한계점이 존재한다. 연구에서는 환자의 최초 심전도와 제일 마지막 심전도에서 특징 값들을 추출하여 간단한 기계 학습 모델인 로지스틱 회귀 모델에 입력하였다. 이러한 접근 방식은 심전도의 시간에 따른 변화 정보가 충분히 반영되기 어렵고, 선택하는 특징에 따라 성능 변동 가능성이 있다는 한계가 존재한다.

본 연구에서는 좌심실 박출률 감소를 보다 정확하게 예측할 수 있는 종단 심전도 기반 예측 프레임워크를 개발하였다. 이는 기존 연구의 한계점을 극복하고, 심전도에 내재된 다양한 정보와 개인별 특성 및

시간에 따른 변화 패턴을 효과적으로 활용하는 것을 목표로 한다. 제안하는 프레임워크는 딥러닝 기술과 종단 심전도 데이터를 결합하여, 심전도의 복잡한 파형과 패턴을 충분히 반영하면서도 특징 추출 과정에서의 정보 손실을 최소화할 수 있도록 설계되었다.

개발된 프레임워크의 주요 장점은 다음과 같다. 첫째, 다중 방문 환자의 원본 심전도 신호를 그대로 입력으로 사용함으로써 특징 추출이 불필요하며 원본 데이터의 풍부한 정보를 최대한 활용할 수 있다. 둘째, 이전 심전도 데이터와 짝을 이루는 심초음파 정보가 없는 경우에도 적용 가능하도록 설계되어, 실제 임상 현장에서의 활용도를 높였다. 셋째, 환자 별 시간에 따른 변화 패턴을 학습함으로써 예측 정확도가 향상 가능하다.

제안하는 프레임워크는 개인별 특성과 시간에 따른 변화 패턴을 학습함으로써, 예측의 정확도와 개인 맞춤화 수준을 향상시킬 수 있을 것으로 기대된다.

제 2 장 연구방법

2.1 연구 대상자

본 연구는 단일 기관에서 수행한 후향적 코호트 연구이다. 서울대학교병원 기관생명윤리위원회의 승인을 받았으며 익명화된 데이터를 사용하는 연구의 특성상 피험자의 동의 절차는 면제되었다. 연구 대상자의 등록 흐름은 Figure 1 에 나타내었다. 2003 년 1 월부터 2022 년 7 월까지 서울대학교병원에서 심초음파 검사를 받은 18 세 이상의 환자를 식별했다. 제외 기준은 다음과 같다: (1) 심초음파 검사결과지 없음($n=29,060$), (2) LVEF 값 누락($n=3,514$), (3) LVEF 값 이상치 ($LVEF < 10$ 또는 $LVEF \geq 80$)($n=802$), (4) 심전도 기록 없음($n=31,104$), (5) 마지막 심초음파 이전 1 년 내에 심전도 기록 존재 없음($n=145,043$), (6) 마지막 심초음파 기준 3 년 전의 심전도 기록 존재 없음($n=144,103$), (7) 10 초 미만인 심전도 존재하거나, 심전도 신호 완전히 누락($n=28$). 최종적으로 환자 별 마지막 심초음파 이전 1 년 내 심전도와 3 년 전 심전도가 존재하는 32,016 명의 환자를 분석했다. 연구 대상자는 대조군($LVEF \geq 50\%$, $n=30,233$)과 LVEF 감소군($LVEF < 50\%$, $n=1,783$)으로 나누었으며 LVEF 감소군의 기준은 클래스 불균형을 고려하여 $LVEF < 50\%$ 을 기준으로 삼았다.

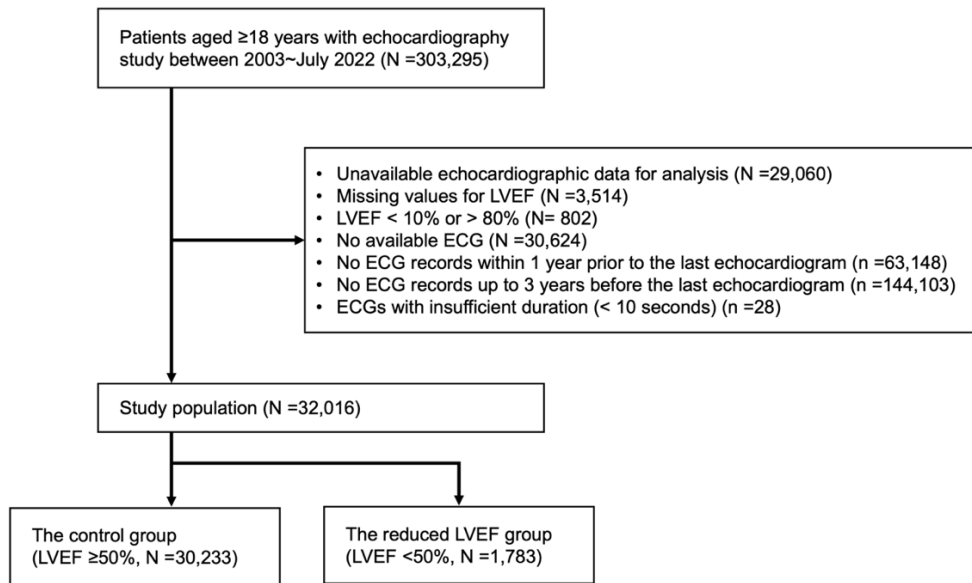


Figure 1. Cohort flow chart

Abbreviations: LVEF, Left Ventricular Ejection Fraction; ECG, Electrocardiogram

환자별로 여러 심초음파 기록과 심전도 기록이 존재할 수 있다. 각 환자의 가장 마지막 심초음파 기록에서의 LVEF 이상 여부를 예측 대상으로 설정하고, 해당 시점을 index date 로 정의하였다(Figure 2). 환자 간 심전도와 심초음파 사이 기간을 통일하기 위해 다음 두 가지 조건을 만족하는 환자들을 선별하였다: (a) index date 를 기준으로 1 년 이내에 측정된 심전도 데이터가 존재하는 경우, (b) index date 를 기준으로 3 년 전에 측정된 심전도 데이터가 존재하는 경우. 선별된 환자들에 대해 (a)와 (b) 사이에 측정된 모든 심전도 데이터를 모델의 입력으로 사용하였다. 즉, index date 기준 1 년 이내에 심전도가 존재하지 않는 환자와 index date 연도 기준 3 년 전 연도에 심전도가 존재하지 않는 환자는 분석에서 제외하였고, 과거 3 년간의 심전도 정보를 바탕으로 LVEF 이상 여부를 예측하고자 하였다.

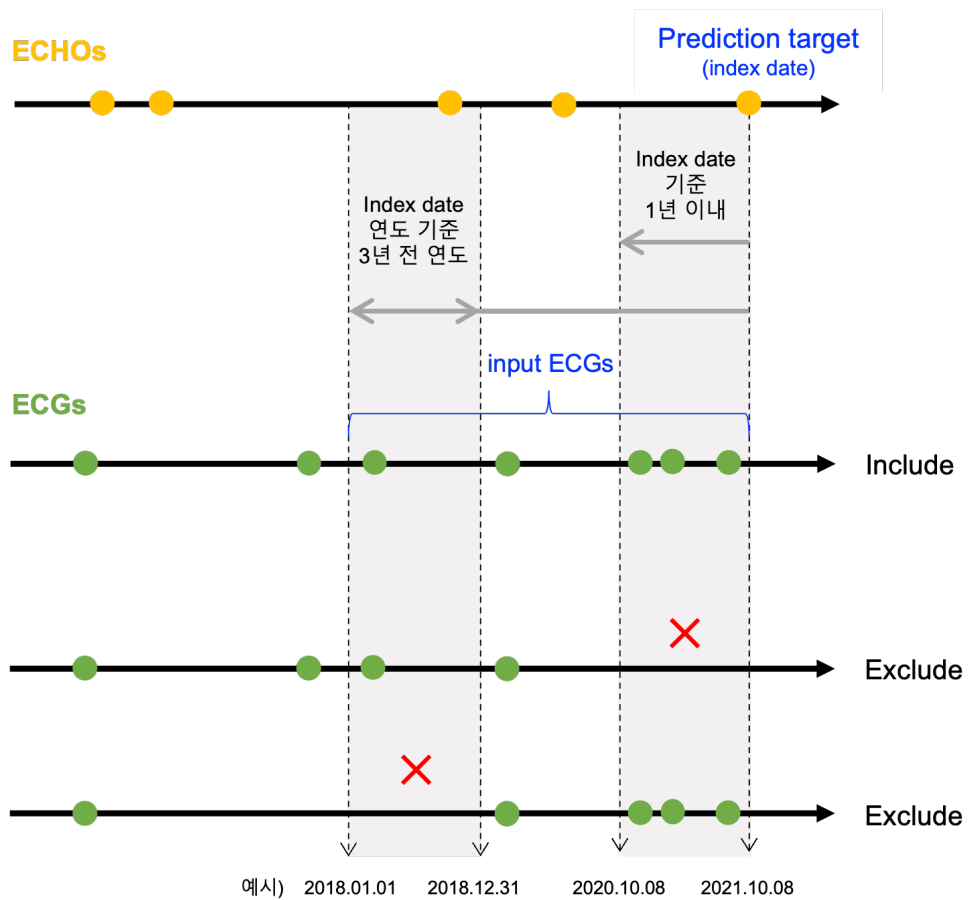


Figure 2. Data selection process example

Abbreviations: ECHO, Echocardiogram; ECG, Electrocardiogram

2.2 데이터 수집

환자의 기본 임상정보 및 심초음파 검사 기록지 데이터는 서울대학교병원 환자 연구 환경 시스템(SUPREME)에서 추출했으며, 심전도 데이터는 MUSE Cardiology Information System (GE Healthcare, WI, USA)에서 추출하였다.

LVEF 값은 순환기 전문의가 확인한 심초음파 검사 기록지 데이터로부터 추출하였다.

원시 심전도 신호 데이터는 XML 파일 내에 base64로 인코딩되어 저장되어 있어 Python 스크립트를 사용하여 숫자 배열로 디코딩 후 사용하였다. 신호는 I, II, III 및 V1-V6의 8개 리드로 구성 되어있고, 기존 리드의 선형 결합을 통해 III, aVR, aVL, aVF의 4가지 리드 신호를 추가로 생성하였다. 500Hz의 샘플링 주파수로 10초 동안 기록된 하나의 심전도 신호는 12x5000의 크기를 나타내며 모델 입력 시 계산 효율성을 위해 샘플링 주파수를 250Hz로 다운 샘플링하여 12x2500의 크기로 입력하였다. Figure 3은 심초음파 검사 기록지에서의 LVEF 값과 12-lead의 원시 심전도 데이터 예시를 나타낸다.

(a)

2017-01-03	
BP: 95 / 69	
HR: 85	
* Final Impression	
1. Normal LV cavity size and systolic function; Calculated EF : 56%	
2. No regional wall motion abnormality	
3. Normal LV wall thickness	
4. Mild AR, Mild to moderate MR	
5. Mild to moderate TR with estimated PASP of 31mmHg	
* Measurement	
LVIDd	47 mm [34-54 mm]
LVIDs	31 mm [23-38 mm]
EF (calculated)	56 % [55-70 %]
IVSd	9 mm [8-11.5 mm]
LVPWd	8 mm [8-11.5 mm]
LA size	46 mm [23-40 mm]
Ao diameter	33 mm [22-33 mm]
Mitral inflow E	1.10 m/s [.5-1.2 m/s]
Mitral inflow DT	140 mSec [150-240 mSec]

(b)

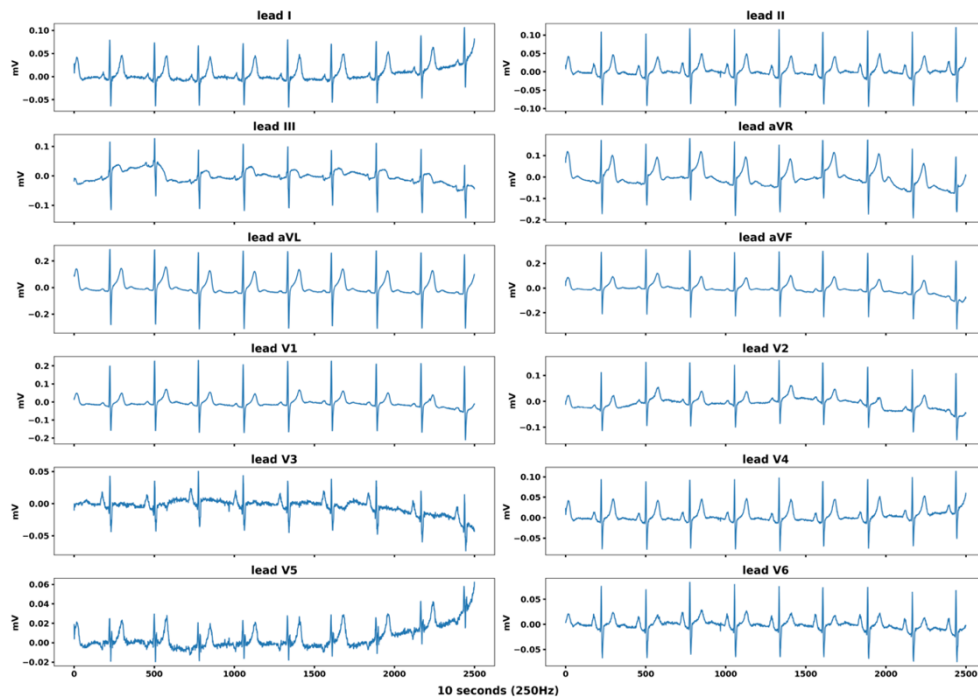


Figure 3. Example of ECHO report and ECG signal (a) Example of ECHO report (b) Example of 12-lead ECG signal
Abbreviations: ECHO, Echocardiogram; ECG, Electrocardiogram

2.3 프레임워크 개발

본 연구에서는 환자의 종단 데이터를 효과적으로 활용하기 위해 가중 합(weighted sum) 기반의 종단 심전도 프레임워크를 제안한다. 이 프레임워크는 다양한 모델 아키텍처와 다양한 태스크에 적용할 수 있으며, 여러 시점의 원시 신호 데이터의 정보를 복잡한 처리나 손실 없이 통합할 수 있다는 장점이 있다. Figure 4는 제안하는 프레임워크의 전체적인 모식도이다. 입력 데이터는 환자의 방문 시점 별 원시 심전도 신호와 각 심전도 측정일과 환자의 마지막 심초음파 측정일 사이의 시간 차(time delta)이다. 딥러닝 모델은 심전도 신호를 처리할 수 있는 구조라면 모두 사용 가능하다. 본 연구에서는 심전도 신호 처리에 보편적으로 사용되는 CNN 기반 모델과 Transformer 기반 모델을 분석에 사용하였으며, 모델에 대한 자세한 내용은 2.5 모델학습 및 평가과정에 작성하였다.

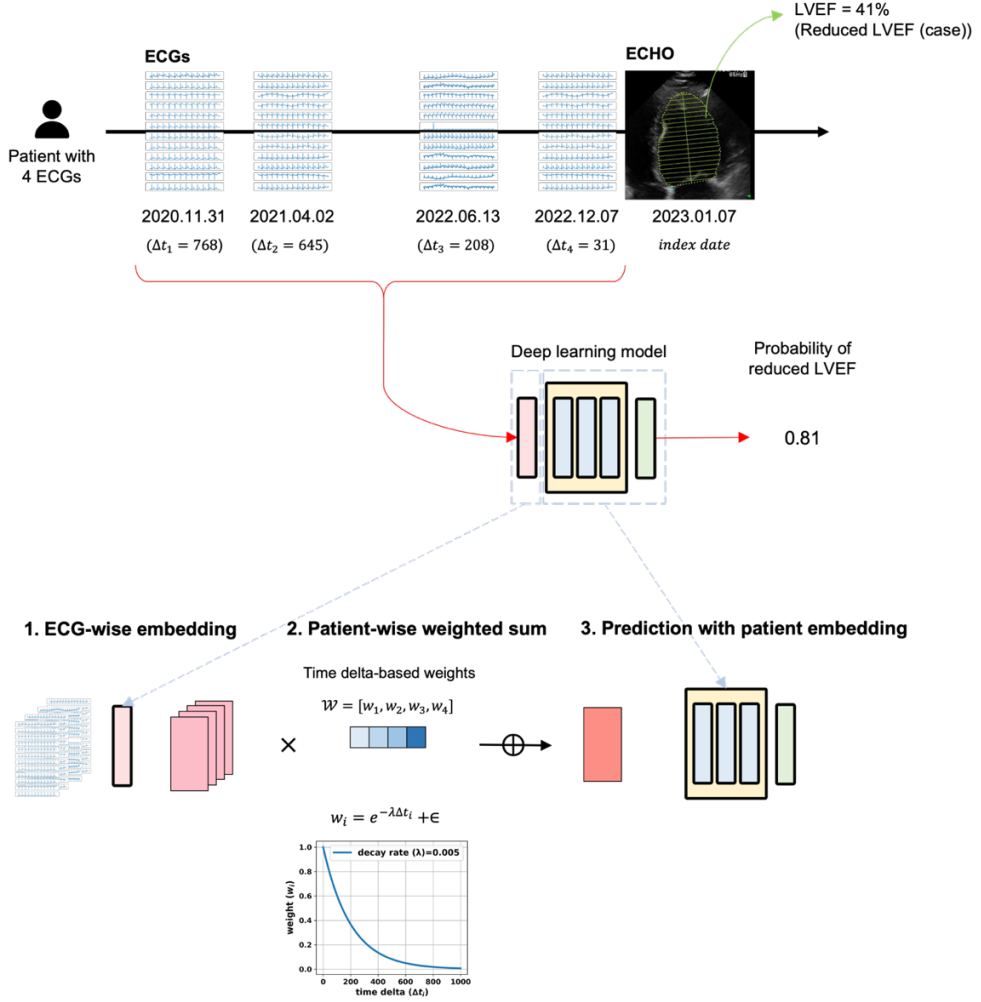


Figure 4. Overview of the Longitudinal ECG framework

Abbreviations: ECHO, Echocardiogram; ECG, Electrocardiogram; LVEF, Left Ventricular Ejection Fraction

프레임워크의 구성 요소는 다음과 같다.

1. **ECG-wise embedding**: 환자의 방문 시점 별 심전도 데이터는 딥러닝 모델 구조의 초반 layer 에 독립적으로 통과한다. ‘초반 layer’란 Transformer 기반의 모델에서의 경우 Patch embedding layer 를 뜻하며, CNN 기반의 모델에서는 반복되는 convolution block 의 집합인 Stage 가 나오기 전 ‘Stem’ 부분을 뜻한다. 입력 데이터의 차원을 축소하고 초기 특징 추출을 수행하는 레이어인 Patch embedding layer 혹은 Stem 부분을 방문 시점 별 심전도가 개별적으로 거치는 것이다. Stem 을 구성하는 convolution layer 가 2 개 이상인 경우 개별 심전도가 통과할 layer 의 수를 'weighted summation depth'이라는 하이퍼파라미터로 정의하였다. Figure 5(a)는 Stem 이 3 개의 convolution layer 로 구성된 ECGNet 의 구조이고, Figure 5(b)와 5(c)는 각각 weighted summation depth 을 1 과 3 으로 설정했을 때의 데이터 처리 과정을 나타낸 것이다. Weighted summation depth 가 클수록 개별 심전도가 더욱 추상화된 형태로 가중 합산 된다.

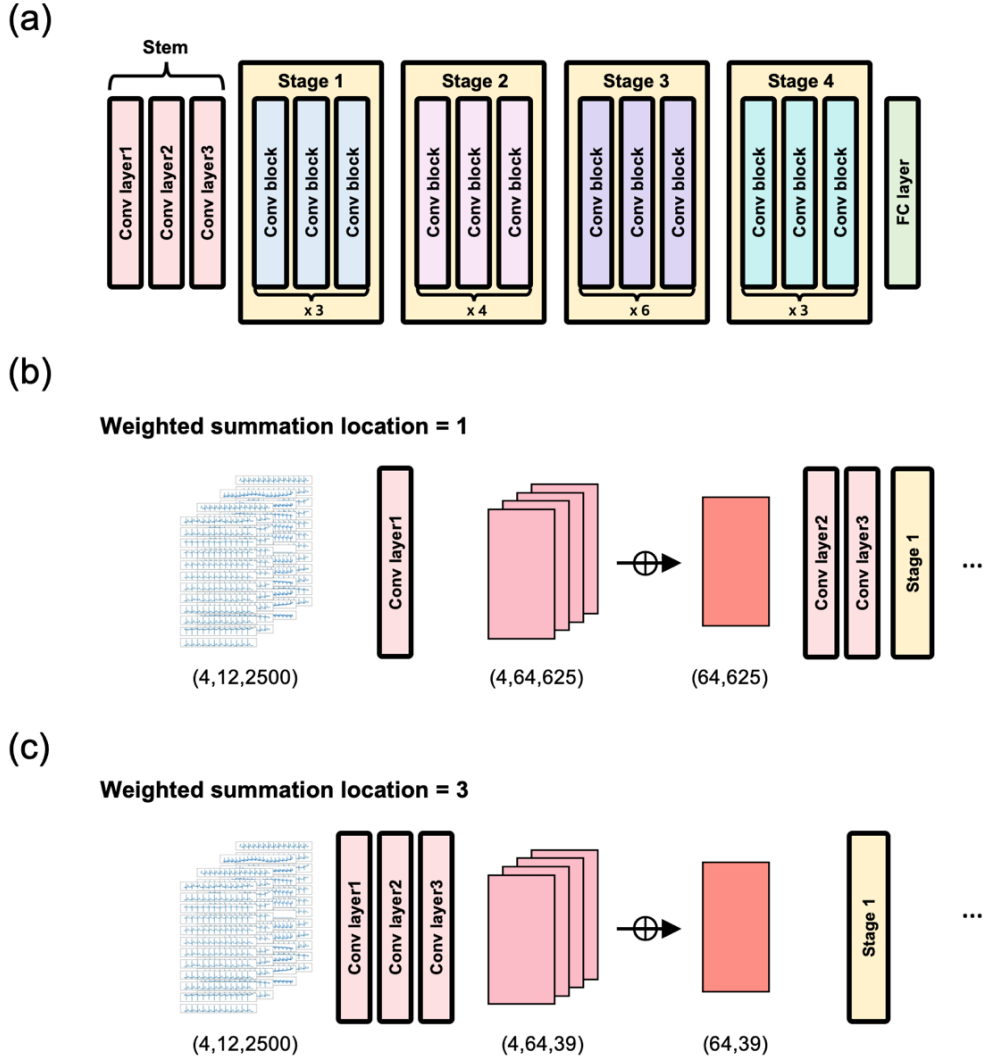


Figure 5. Weighted summation locations in ECGNet (a) Overview of ECGNet architecture (b) Calculation process when weighted summation location is 1 (c) Calculation process when weighted summation location is 3

Abbreviations: Conv, Convolution; FC, Fully Connected

ECG-wise embedding 을 수행하기 전에 환자 간 심전도 데이터 수가 상이하다는 점을 고려하여 입력 데이터의 크기를 일관되게 유지하기 위한 방법이 적용되었다. 최대 심전도 개수를 나타내는 파라미터 max length 를 설정하고, 자연어처리 모델에서 가변 길이 문장을 처리하는 방식과 유사하게 접근하였다. max length 보다 적은 수의 심전도를 가진 환자의 경우 max length 와 실제 심전도 수의 차이만큼 제로 패딩을 적용하여 입력 크기를 통일하였다. max length 를 초과하는 심전도를 가진 환자의 경우 균등한 간격으로 max length 개의 데이터를 샘플링하여 사용하였다. Figure 6 은 11 개의 심전도를 갖는 환자에서 5 개를 샘플링할 때의 예시를 나타낸 것이다. 처음과 끝 심전도는 포함하고 그 사이를 균등한 간격으로 나누어 해당 간격 값과 제일 가까운 데이터를 선택하였다. 결과적으로, 모델에 입력되는 데이터의 차원은(batch size, max length, 12, 2500)로 표준화되었다. 본 연구에서는 전체 환자의 심전도 개수 분포를 분석하여, 90% 해당하는 15 를 max length 값으로 설정하였다.

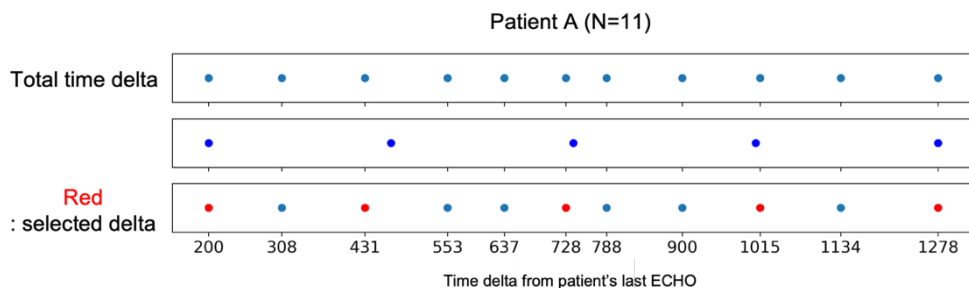


Figure 6. Example of equally sampling strategy

Abbreviations: ECHO, Echocardiogram

2. Patient-wise weighted sum: 1 단계에서 얻어진 방문 시점 별 embedding 은 가중 합을 통해 통합되어 최종 예측을 수행한다. 가중치는 시점 별 중요도를 반영하며, 이를 통해 종단 데이터의 정보를 효과적으로 통합할 수 있다. 본 연구에서는 지수 함수를 가중치 함수로 채택하였다. 지수 가중 함수는 index date 와 각 심전도의 측정일 사이의 시간 차이(Δt)에 기반하여 가중치를 할당하며, Δt 가 작을수록 높은 가중치를 부여한다. 이를 통해 최신 데이터의 중요성을 강조하면서도 과거 데이터의 정보도 활용할 수 있게 한다. 가중치는 수식 2 를 통해 계산되었다.

$$w_{ik} \begin{cases} e^{-\lambda \Delta t_{ik}} + \epsilon & \text{if } k \neq n_i \\ 1 & \text{if } k = n_i \end{cases} \quad (2)$$

w_{ik} 는 환자 i 의 k 번째 심전도 데이터에 적용된 가중치를 나타내며 0에서 1 사이의 값이다. n_i 는 환자 i 의 심전도 개수를 나타내며, λ 는 지수함수의 감쇠율(decay rate), Δt_{ik} 는 환자 i 의 index date 와 k 번째 심전도 측정일 사이의 시간 차이(일), ϵ 은 계산의 안정성을 보장하기 위해 추가된 매우 작은 상수(1×10^{-10})이다. $k = n_i$ 일 경우, 즉 환자의 마지막 심전도의 경우 index date 와 가장 근접하여 가장 높은 정보 가치를 지니므로 최대 가중치인 1 을 가해준다. 그보다 앞선 심전도들에 대해서는 지수함수를 따르는 가중치를 곱해준다.

Figure 7 은 다양한 감쇠율에 따른 가중 함수를 시각화 한 것이다. Figure 7(a)는 $\lambda = 0.001$ 일 때의 상대적으로 완만한 감쇠율을, Figure 7(c)는 $\lambda = 0.05$ 일 때의 급격한 감쇠를 보여준다. 낮은 감쇠율에서는 지수함수의 기울기가 완만하여 Δt 가 큰 과거 데이터에도 상당한 가중치가 부여되어 과거 정보가 더 많이 반영됨을 관찰할 수 있다.

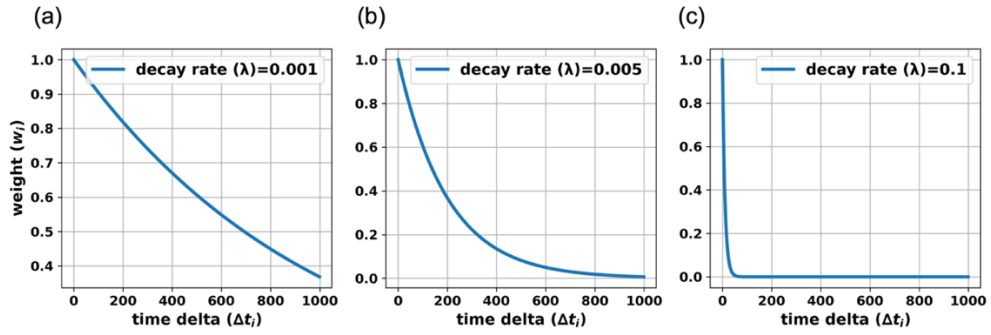


Figure 7. Examples of exponential weighting functions (a) Function with a small gamma value ($\lambda=0.001$) (b) Function with a middle gamma value ($\lambda=0.005$) (c) Function with a large gamma value ($\lambda=0.1$)

3. Prediction with patient embedding: 2 단계에서 가중 합을 거치면 gradient explosion 을 막기 위해 환자 임베딩 별 평균, 표준편차를 활용하여 정규화를 시켜준다. 이렇게 생성된 patient embedding 은 환자의 종단 심전도 정보를 압축적으로 표현한다. patient embedding 이 모델의 후반부 레이어에 입력되어 최종 예측을 수행한다.

2.4 모델 학습 및 평가과정

본 연구에서는 공정한 성능 평가를 위해 Time hold-out split 을 진행하여 2021 년 7 월 1 일 이전의 데이터($n=27,492$)를 모델 개발을 위해 사용하였으며, 2021 년 7 월 1 일 이후의 데이터($n=7,224$)는 검증용 테스트 셋으로 활용하였다.

모델 구조에 상관없이 longitudinal ECG 프레임워크가 우수한 성능을 보이는지 확인하기 위해 CNN 기반 ResNet[31] 계열 모델들, ECGNet, Transformer 기반 Vision Transformer (ViT)[32] 계열 모델을 분석에 사용하였다. ResNet 은 잔차 연결(residual connection)을 도입하여 깊은 layers 의 학습을 용이하게 한 CNN 모델로, 심전도를 비롯한 생체신호 분석 연구에서 기준 모델(baseline)로 널리 활용되고 있다. ECGNet 은 VGGNet 과 ResNet 구조를 기반으로 설계된 생체신호 처리에 특화된 nEMGNet[33]을 참고하여 설계하였으며, 일부 convolution layer 의 커널 크기를 심전도 신호의 입력 형태에 맞게 조정하였다. ViT 는 Transformer 구조를 이미지 분류 문제에 적용한 모델로, 이미지를 패치로 분할하여 Transformer 인코더에 입력하는 방식을 채택하고 있다. 최근 연구에 따르면 ViT 의 패치 분할 및 토큰화 방식이 심전도와 같은 시계열 데이터 처리에도 효과적인 것으로 나타났기 때문에 선정하였다. 다양한 모델 크기에 따른 성능 비교를 위해 ViT-Tiny (9M), ResNet18 (9M), ViT-Small (25M), ResNet50 (25M)을 실험에 포함하였다.

모델의 최적 하이퍼파라미터 선정을 위하여 개발 셋을 랜덤하게 훈련 셋과 검증 셋으로 88:12 의 비율로 분할하였으며(훈련 셋 $n=21,816$, 검증 셋 $n=2,976$), 검증 셋에서의 F1 score 가 가장 높은 하이퍼파라미터 조합을 최종 파라미터로 선정하였다.

하이퍼파라미터로는 학습률, 배치 크기, 감쇠율(decay rate), Weighted summation depth, 패치 길이를 고려하였으며, Weighted summation depth 은 ECGNet 에서만 고려하였고, 패치 길이는 ViT 계열 모델에서만 고려하였다. Hyperparameter tuning 관련한 정보는 Table 1 에 나타내었다. 70 epoch 으로 학습을 진행하였고, 과적합을 방지하기 위해 valid loss 를 기준으로 7 epoch 이상 하락이 없을 경우 early stopping 을 수행하였다. 이중분류 문제이므로 손실함수는 Binary cross entropy loss 를 사용하였다.

Table 1. Hyperparameter tuning set

Hyperparameter	Tuning set
Learning rate	0.001, 0.0005
Batch size	32, 64, 256, 512
Decay rate	0.001,0.005,0.020
Weighted summation depth (ECGNet)	1, 2, 3
Patch length (sec) (ViT)	0.1, 0.2, 0.4, 0.6, 0.8

2.5 통계 분석

성능 평가를 위해 사용된 진단 지표들은 다음과 같다: 수신기 조작 특성 곡선 하의 면적(Area under the receiver operating characteristic curve; AUROC), 정밀도-재현율 곡선 하의 면적(Area under the precision-recall curve; AUPRC), F1 점수, 양성 예측도(Positive Predictive Value; PPV), 음성 예측도(Negative Predictive Value; NPV). 임계 값에 의존하는 지표들의 최적의 임계점(threshold)을 결정하기 위해서 유덴 J-지수(Youden J-index)[34]를 사용하였다(수식 3). 최적의 임계점을 통해 true positive (TP), false positive (FP), false negative (FN), true negative (TN)를 구하고 PPV, NPV, F1 점수를 계산하였다(수식 4,5,6). AUROC 는 들롱 검정(DeLong test)을 사용하여 비교하였다. 양측 검정에서 P 값이 .05 미만일 경우 귀무 가설을 기각하였다. 모든 통계 분석은 R 소프트웨어와 파이썬을 사용하여 수행되었다.

$$Youden's J statistic = Sensitivity + Specificity - 1 \quad (3)$$

$$PPV = \frac{TP}{TP+FP} \quad (4)$$

$$NPV = \frac{TN}{TN+FN} \quad (5)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision+Recall}, Precision = \frac{TP}{TP+FP}, Recall = \frac{TP}{TP+FN} \quad (6)$$

제 3 장 연구결과

3.1 데이터 구성

총 32,036 명의 환자로 부터 얻은 211,592 개의 심전도 데이터가 분석에 사용되었다. 대조군은 30,233 명의 환자로 부터 얻은 183,970 개의 심전도로 구성되었으며, 사례군(좌심실 박출률 감소군)은 1,783 명의 환자로 부터 얻은 27,622 개의 심전도로 이루어졌다. 환자들의 인구통계학적 특성과 심전도 특성은 Table 2 에 나와있다.

인구통계학적 특성을 보면, 사례군의 평균 연령이 대조군보다 유의하게 높았고(73.0 세 vs. 63.0 세, $P < 0.001$), 남성의 비율 또한 사례군에서 유의하게 높았다(67.7% vs. 57.4%, $P < 0.001$). 신장과 체질량지수는 사례군에서 대조군보다 유의하게 컸으나(164.1cm vs. 163.0cm, $P = 0.027$; $24.0\text{kg}/\text{m}^2$ vs. $24.2\text{kg}/\text{m}^2$, $P = 0.005$), 체중(64.4kg vs. 64.2kg, $P = 0.457$)은 유의한 차이가 없었다.

심전도 특성의 경우 사례군은 대조군에 비해 심전도 시행 횟수가 많았고(10.0 회 vs. 4.0 회, $P < 0.001$), 심박수(77.0 회/분 vs. 70.0 회/분, $P < 0.001$), QRS 간격(106.0ms vs. 94.0ms, $P < 0.001$), QT 간격(416.0ms vs. 406.0ms, $P < 0.001$)이 유의하게 증가하였다. 또한 사례군에서 낮은 R 축(-22.0 도 vs. 39.0 도, $P < 0.001$)과 높은 T 축(87.0 도 vs. 45.0 도, $P < 0.001$)이 관찰되었고, 동성 리듬(42.0% vs. 57.4%, $P < 0.001$), 동성 서맥(6.8% vs. 15.8%, $P < 0.001$)의 비율이 감소한 반면, 동성 빈맥(6.4% vs. 3.7%, $P < 0.001$), 심방세동(22.4% vs. 12.7%, $P < 0.001$), 심방조동(2.7% vs. 1.8%, $P < 0.001$), 심방빈맥(0.1% vs. 0.008%, $P < 0.001$)의 비율은

증가하였다. 우각차단(9.1% vs. 8.7%, $P=0.011$)과 좌각차단(5.4% vs. 1.0%, $P<0.001$)의 유병률도 사례군에서 유의하게 높았다(Table 2).

Table 2. Baseline characteristics

	Control group (EF ≥50%)	Case group (EF <50%)	P-value
Patients (n)	30,233	1,783	-
ECGs (n)	183,708	27,579	-
<i>Demographics</i>			
Age (yr)	63.0 (55.0-73.0)	73.0 (64.0-79.0)	<0.001
Male (%)	17,356 (57.4)	1,207 (67.7)	<0.001
Height (cm)	163.0 (156.0-169.7)	164.1 (157.3-169.4)	0.027
Body weight (kg)	64.2 (56.4-72.3)	64.4 (56.275-71.9)	0.457
BMI (kg/m ²)	24.2 (22.2-26.3)	24.0 (22.0-26.1)	0.005
<i>Electrocardiogram</i>			
Number of ECG	4.0 (3.0-6.0)	10.0 (6.0-20.0)	<0.001
Heart rate (/min)	70.0 (61.0-81.0)	77.0 (66.0-91.0)	<0.001
QRS duration (ms)	94.0 (86.0-104.0)	106.0 (94.0-134.0)	<0.001
QT interval (ms)	406.0 (380.0-434.0)	416.0 (380.0-460.0)	<0.001
R-axis (degree)	39.0 (8.0-66.0)	22.0 (-22.0-63.0)	<0.001
T-axis (degree)	45.0 (25.0-66.0)	87.0 (35.0-128.0)	<0.001
Q onset	219.0 (214.0-224.0)	216.0 (211.0-220.0)	<0.001
Q offset	266.0 (261.0-271.0)	269.0 (264.0-280.0)	<0.001
Sinus rhythm (%)	105,473 (57.4)	11,576 (42.0)	<0.001
Sinus tachycardia (%)	6,849 (3.7)	1,771 (6.4)	<0.001
Sinus bradycardia (%)	29,079 (15.8)	1,872 (6.8)	<0.001
Atrial fibrillation (%)	23,247 (12.7)	6,167 (22.4)	<0.001
Atrial flutter (%)	3,335 (1.8)	749 (2.7)	<0.001
Atrial tachycardia (%)	91 (0.05)	35 (0.1)	<0.001
RBBB (%)	15,948 (8.7)	2,523 (9.1)	0.011

LBBB (%)	1,858 (1.0)	1,479 (5.4)	<0.001
----------	-------------	-------------	--------

Data are N (%) or median (interquartile range). Abbreviations: ECG, Electrocardiogram; BMI, Body mass index.

3.2 Cross-sectional vs. Longitudinal 성능 비교

본 연구에서는 다양한 딥러닝 모델의 성능을 파라미터 크기와 학습 방법에 따라 비교 분석하였다(Table 3). ResNet, ViT, ECGNet 모델을 사용하여 제일 최신 시점의 심전도만 학습에 활용하는 cross-sectional (cross.) 방법과 환자의 모든 심전도 데이터를 활용하는 longitudinal (long.) 방법의 성능을 비교하였다.

전반적으로 cross. 방법보다 long. 방법의 성능이 우수하였다. AUROC 지표에서는 모든 모델에서 long. 방법이 cross. 방법보다 높은 성능을 보였다(AUROC, ResNet18 0.922 vs 0.923; ViT-T 0.900 vs 0.904; ResNet50 0.906 vs 0.919; ViT-S 0.895 vs 0.907; ECGNet 0.925 vs 0.925). 25M 모델 중 ResNet50, ViT-S 는 AUROC 가 통계적으로 유의미하게 long.이 우수한 성능을 냈다(ResNet50 $p=0.015$, ViT-S $p=0.037$). F1 점수와 PPV 의 경우, ResNet50 을 제외한 모든 모델에서 long. 방법이 더 높은 성능을 나타냈다(F1 score: ResNet18 0.456 vs 0.506; ViT-T 0.434 vs 0.437; ResNet50 0.473 vs 0.472; ViT-S 0.458 vs 0.463; ECGNet 0.499 vs 0.532, PPV: ResNet18 0.314 vs 0.380; ViT-T 0.302 vs 0.308; ResNet50 0.342 vs 0.338; ViT-S 0.326 vs 0.335; ECGNet 0.364 vs 0.410).

ECGNet (long.)은 AUROC (0.925), F1 (0.532), AUPRC (0.572), PPV (0.410) 4 가지 지표에서 전체 모델 중 가장 높은 성능을 달성했으므로 이후 분석에서 baseline model 로 사용하였다.

Table 3. Performance comparison between longitudinal and cross-sectional settings

Param. size	Model	AUROC	P- value	F1	AUPRC	PPV	NPV
9M	ResNet18 (cross.)	0.922	0.794	0.456	0.544	0.314	0.986
	ResNet18 (long.)	0.923		0.506	0.550	0.380	0.981
	ViT-T (cross.)	0.900	0.541	0.434	0.472	0.302	0.981
	ViT-T (long.)	0.904		0.437	0.425	0.308	0.979
25M	ResNet50 (cross.)	0.906	0.015	0.473	0.467	0.342	0.981
	ResNet50 (long.)	0.919		0.472	0.517	0.338	0.982
	ViT-S (cross.)	0.895	0.037	0.458	0.465	0.326	0.981
	ViT-S (long.)	0.907		0.463	0.468	0.335	0.979
	ECGNet (cross.)	0.925	0.871	0.499	0.533	0.364	0.983
	ECGNet (long.)	0.925		0.532	0.572	0.410	0.981

Bold red values indicate the highest performance, while **blue values** indicate the second highest performance.

3.4 기계학습 모델을 활용한 성능 비교

선행 연구[30]에서는 종단 심전도 데이터를 활용하여 기계학습 모델을 적용하였으며, 가장 최신 시점과 가장 이전 시점의 심전도에서 추출한 특징들을 표 형태로 구성하여 입력하였다. 본 연구의 태스크와는 차이가 있으나, 프레임워크 간 성능 비교를 위해 다양한 기계학습 모델을 사용하여 좌심실 박출률 감소 예측 성능을 분석하였다. 분석에는 선행 연구에서 사용된 Logistic Regression [35]뿐 아니라 또 다른 선형 모델인 Linear Discriminant Analysis (LDA) [36]와 트리 기반 앙상블 모델인 LightGBM [37] Extra Trees Classifier [38]를 사용하였다.

Cross-sectional (cross.) 설정에서는 각 환자의 가장 최근 시점 심전도에서 추출한 15 개의 심전도 특징과 함께 나이, 성별을 포함한 총 17 개의 특징을 입력 변수로 사용하였다. Longitudinal (long.) 설정에서는 각 환자의 첫 번째 시점 심전도에서 추출한 15 개의 심전도 특징, 마지막 시점 심전도에서 추출한 15 개의 특징, 두 시점에서의 나이(2 개), 성별, 그리고 첫 번째와 마지막 심전도 간의 시간 간격을 포함하여 총 34 개의 특징을 입력 변수로 활용하였다.

Table 4 를 보면 long. 성능이 cross.보다 전반적으로 우수한 것으로 나타났다. AUROC 의 경우 모든 모델에서 cross.보다 long.이 높았고(Logistic Regression 0.841 vs. 0.846; LDA 0.835 vs. 0.846; Extra Trees Classifier 0.714 vs. 0.887; LightGBM 0.881 vs. 0.893), Logistic Regression 을 제외한 모든 모델에서 통계적으로 유의미한 차이를 보였다.

F1 score 또한 Extra Trees Classifier 를 제외한 모델들에서 long.의 성능이 높았고(Logistic Regression 0.334 vs. 0.335; LDA 0.349 vs. 0.360; LightGBM 0.367 vs. 0.370) AUPRC 는 LDA 를 제외한

모델들에서 long.의 성능이 높았다(Logistic Regression 0.332 vs. 0.344; LDA 0.323 vs. 0.318; Extra Trees Classifier 0.218 vs. 0.399).

기계학습 모델 중 2 가지 지표(AUROC, F1 score)에서 가장 우수한 성능을 보이고 2 가지 지표(AUPRC, NPV)에서 두 번째로 우수한 성능을 보인 LightGBM long.의 성능과 ECGNet long.의 성능을 비교하면, ECGNet 의 성능이 NPV 를 제외한 모든 지표들에서 우수한 것을 확인할 수 있다(AUROC 0.893 vs. 0.925, p -value < 0.001 ; AUPRC 0.398 vs. 0.532; F1 0.370 vs. 0.572; PPV 0.234 vs. 0.410; NPV 0.986 vs. 0.981) (Table 5).

Table 4. Performance comparison between longitudinal and cross-sectional settings on machine learning models

Model	AUROC	p-value	AUPRC	F1	PPV	NPV
Logistic Regression (cross.)	0.841	0.174	0.332	0.334	0.214	0.979
Logistic Regression (long.)	0.846		0.344	0.335	0.209	0.979
Linear Discriminant Analysis (cross.)	0.835	0.029	0.323	0.349	0.226	0.978
Linear Discriminant Analysis (long.)	0.846		0.318	0.360	0.239	0.979
Extra Trees Classifier (cross.)	0.714	<0.001	0.218	0.352	0.257	0.963
Extra Trees Classifier (long.)	0.887		0.399	0.346	0.208	0.988
LightGBM (cross.)	0.881	<0.001	0.385	0.367	0.236	0.983
LightGBM (long.)	0.893		0.398	0.370	0.234	0.986

Bold red values indicate the highest performance, while **blue values** indicate the second highest performance.

Table 5. Performance comparison between the best performing deep learning and machine learning models in longitudinal settings

Model	AUROC	p-value	AUPRC	F1	PPV	NPV
LightGBM (long.)	0.893	<0.001	0.398	0.37	0.234	0.986
ECGNet (long.)	0.925		0.532	0.572	0.41	0.981

Bold values indicate the highest performance.

3.5 예측 결과에 대한 해석 가능성

본 연구에서는 Expected Gradients 기법 [39]을 활용하여 개별 환자의 다중 방문 데이터에 대한 중요도를 분석하였다. 이 알고리즘을 통해 각 방문 시점의 심전도 신호와 특징들이 모델의 예측에 미치는 영향을 파악할 수 있다. 먼저, 각 방문 시점별로 심전도 신호와 특징들에 대한 Expected Gradient 를 계산하고, 절대값을 합산하여 방문별 중요도를 도출하였다. 이를 통해 환자의 방문 이력 중 어떤 시점이 모델의 예측에 가장 큰 영향을 미쳤는지 확인할 수 있다.

다음으로, 방문 별 중요도가 가장 높게 나타난 시점의 12-lead 심전도 신호에 대해 Integrated Gradients 기법 [40]을 통해 신호 내 중요 구간을 파악하였다. 이는 해당 시점의 심전도 신호 중 모델이 주목한 부분을 시각화 함으로써 좌심실 박출률 감소 예측에 있어 중요한 역할을 하는 신호의 특징을 이해하는 데 도움을 준다 [27].

총 15 개의 방문 데이터가 존재하는 좌심실 박출률 감소 환자에 대해 분석 결과를 확인해보았고, 7 번째 방문의 중요도가 가장 높게 나타났다. 7 번째 방문에서의 ECG feature 들을 살펴보면 Heart rate, QRS duration, QT interval, R axis 가 모두 정상치를 벗어난 것을 확인할 수 있다. 또한 기존에는 보이지 않았던 RBBB 도 함께 나타난 것을 확인하였다(Figure 8 (a)).

다음으로, 방문별 중요도가 가장 높게 나타난 7 번째 방문의 12-lead 심전도 신호에 대해 Integrated Gradients 기법을 통해 신호 내 중요 구간을 파악하였다. Figure 8(b)를 보면, V5 lead 를 비교적 중요시 보는 것을 확인하였고, 12 lead 전반적으로 신호의 QRS 복합파 혹은 P peak, T peak 을 중요시 보는 것을 확인하였다. V5 lead 는 좌심실 활동을 잘 파악할 수 있는 lead 로 알려져 있으며, QRS complex 는 심실의

활성화를 나타내는 지표로 모델이 임상적 정보를 바탕으로 추론을 수행한 것을 확인하였다.

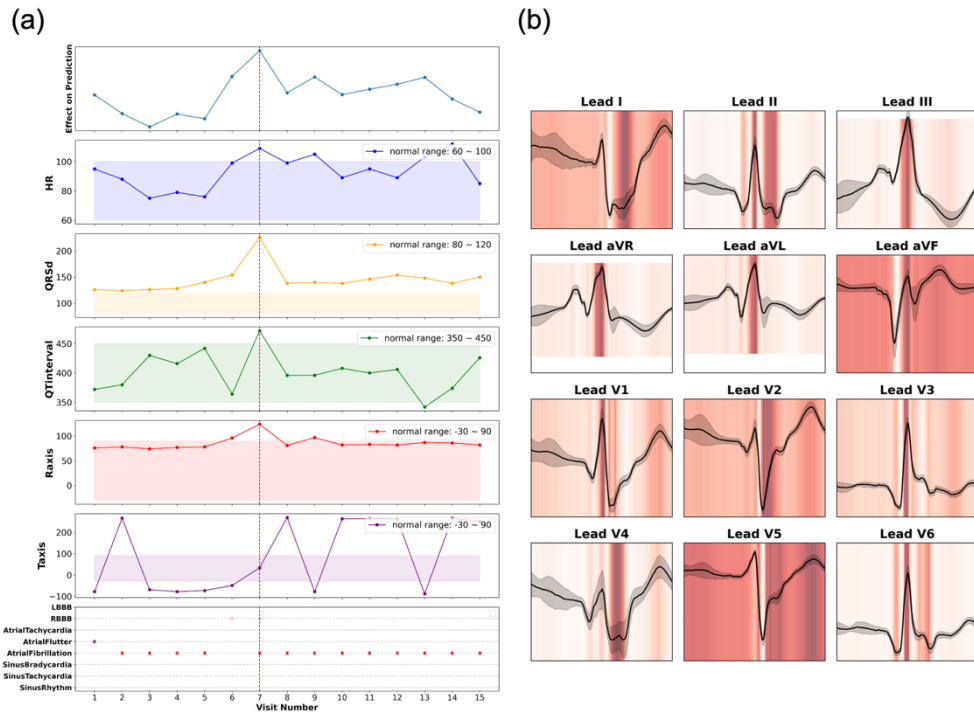


Figure 8. Model explainability (a) Feature importance and ECG feature values across visits (b) Integrated Gradient values of ECG signals at the 7th visit with the highest importance of the Longitudinal ECG framework

Abbreviations: HR, Heart Rate; QRSd, QRS duration; LBBB, Left Bundle Branch Block; RBBB, Right Bundle Branch Block

3.5 심전도 개수에 따른 성능 변화 분석

Figure 9 는 환자별 심전도 검사 횟수가 모델 성능에 미치는 영향을 나타낸다. 모델 성능은 F1 점수와 양성 예측도(PPV)를 통해 평가되었으며, 테스트 세트에서 사용되는 심전도 횟수를 단계적으로 변화시켜 성능 변화를 관찰하였다. 본 실험의 목적은 longitudinal setting 에서 관찰된 성능 향상이 단순히 학습에 사용된 전체 심전도 데이터 양의 증가에 기인한 것이 아님을 입증하는 것이다. 만약 데이터 양의 증가만이 성능 향상의 주된 요인이라면, 학습이 완료된 모델로 테스트 세트 환자에 대해 추론할 때, 환자 별 입력 심전도 횟수에 따른 성능 차이가 미미해야 한다. 분석 결과, F1 점수와 PPV 모두 심전도 검사 횟수가 증가함에 따라 지속적으로 상승하는 경향을 보였다. 특히 4 회 이상의 심전도 검사를 사용한 경우부터 성능이 급격히 향상되었으며, 이후에도 점진적인 증가 추세가 관찰되었다. 이러한 결과는 본 모델이 단순히 데이터의 양적 증가를 넘어서, 환자의 시간에 따른 심전도 변화 패턴을 효과적으로 학습하고 활용하고 있음을 시사한다.

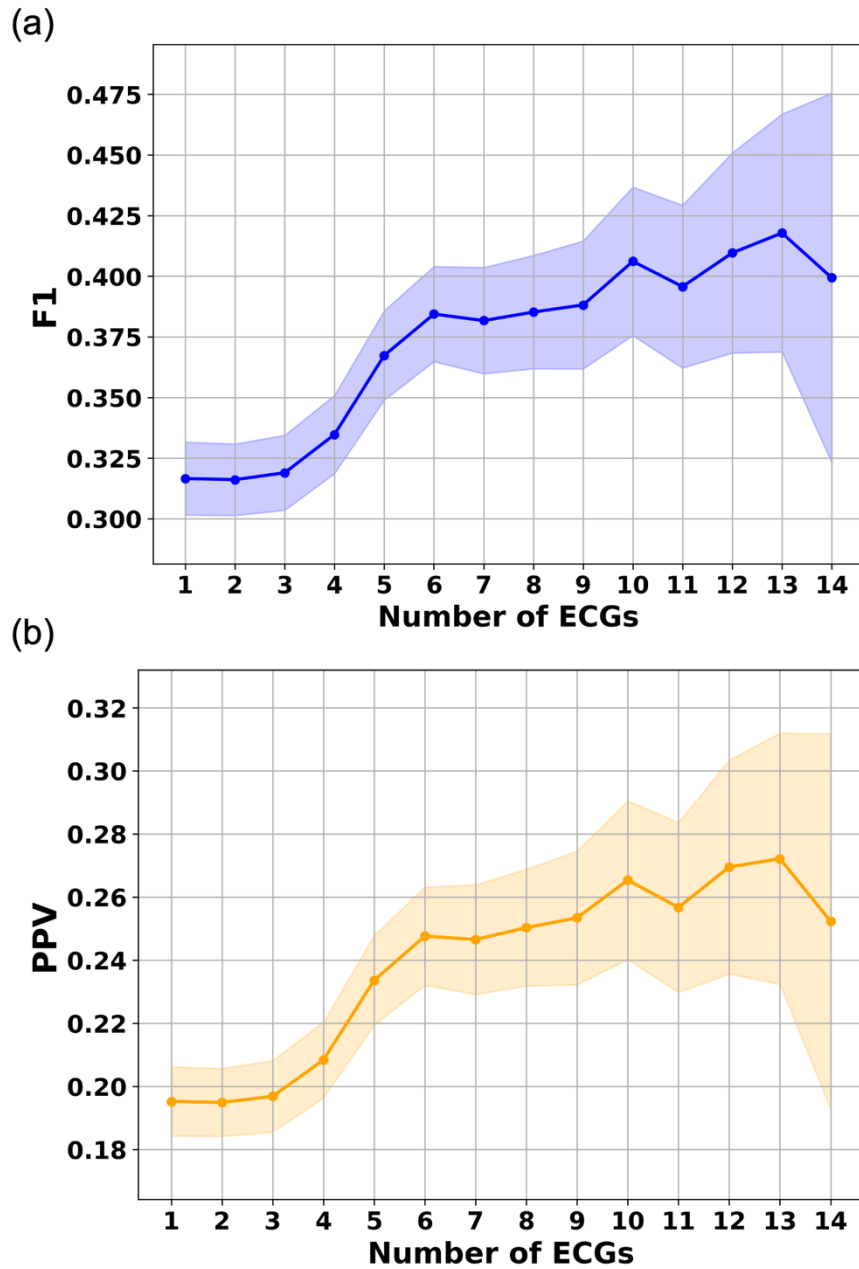


Figure 9. The impact of the number of ECGs on model performance (a) F1 score plotted against the number of ECG used for model evaluation on the test set (b) PPV plotted against the number of ECG used for model evaluation on the test set
Abbreviations: ECG, Electrocardiogram; PPV, Positive Predictive Value
Shaded areas indicated the standard deviation of the corresponding performance measures

3.6 파라미터에 따른 성능 분석

제안된 종단 심전도 분석 프레임워크의 성능에 영향을 미치는 주요 하이퍼파라미터인 감쇠율(Decay rate), 가중 합산 깊이(Weighted summation depth), 샘플링 방법(Sampling strategy), 패딩 방법(Padding strategy)에 따른 성능 변화를 분석하였다(Table 6).

감쇠율은 시간 차이에 따른 가중치 값을 결정하는 파라미터로, 최적 값 탐색을 위해 다양한 값들을 실험하였다. 그 결과, 0.005까지는 성능이 향상하다가 그 이후부터는 다시 감소하는 것을 확인하였다. 이는 과거의 데이터에 너무 큰 가중치가 부여되면 최신의 데이터가 가려져 성능에 방해되지만, 적절한 비율로 과거 데이터 정보가 반영된다면 성능 향상에 도움되는 것을 나타낸다. 한편, 모든 시점에 동일한 가중치 1을 부여하는 방식(No weight)과 비교했을 때 대부분의 감쇠율 설정에서 더 나은 성능을 보여 가중치 할당의 필요성과 효과를 입증하였다.

가중 합산 깊이는 개별 심전도 데이터가 독립적으로 통과하는 초기 컨볼루션 층의 수를 결정하는 파라미터이다. 실험 결과, 3 번째 컨볼루션 층 이후에 가중 합산을 수행하는 것이 가장 우수한 성능을 달성하였다(F1 0.532; PPV 0.410). 이는 데이터의 추상화 수준과 가중 합산 시점 간의 관계가 성능에 영향을 미칠 수 있음을 나타낸다. 적절한 수준의 특징 추출 이후에 가중 합산을 수행하는 것이 개별 심전도의 정보를 효과적으로 통합하는 데 도움이 되는 것으로 해석할 수 있다.

샘플링 방법(Sampling strategy)은 환자의 심전도 개수가 설정된 최대 개수(Max length)를 초과할 때 어떤 방식으로 심전도를 선택할지를 결정하는 파라미터이다. 논문에서 채택한 기본 방식인 균일 간격 샘플링(Equality sample)과 제일 최신의 max length 개를 선택하는 최신 샘플링(Latest sample) 두 가지 방법을 비교하였다. 실험 결과, 균일

간격 샘플링 방식이 더 우수한 성능을 보였다(F1 0.532; PPV 0.410). 이는 환자의 전반적인 심전도 변화 패턴을 포착하는 데 있어 시간적으로 고르게 분포된 샘플이 더 효과적임을 시사한다.

패딩 방법(Padding strategy)은 환자의 심전도 개수가 Max length 이하일 때 심전도 개수-max length 만큼의 공란을 어떠한 방식으로 채울지 나타내는 파라미터이다. 논문에서 채택한 기본 세팅인 제로 패딩과 환자에게 존재하는 데이터를 반복하여 공란을 채우는 Amplification 방법을 비교하였다. 실험 결과 zero padding 이 더욱 우수한 성능을 보였고, 환자 별 데이터 개수를 맞출 때 데이터가 없는 부분에는 단순히 0 을 입력하여도 충분히 좋은 결과를 보임을 확인하였다.

Table 6. Performance per hyperparameter value

Parameter	Value	F1	PPV
Decay rate	No weight	0.476	0.338
	0.001	0.467	0.338
	0.002	0.477	0.333
	0.005	0.532	0.410
	0.01	0.480	0.339
	0.02	0.505	0.371
	0.1	0.468	0.328
	0.3	0.458	0.317
Weighted summation depth	No embedding	0.454	0.322
	1	0.466	0.331
	2	0.498	0.357
	3	0.532	0.410
Sampling strategy	Equally sample	0.532	0.410
	Latest sample	0.519	0.391
Padding strategy	Zero padding	0.532	0.410
	Amplification	0.512	0.396

Bold values indicate the highest performance.

3.7 클래스 불균형: 데이터 처리 기법의 영향 분석

의료 분야에서 딥러닝 적용 시 대조군이 사례군보다 많은 클래스 불균형 문제는 흔히 발생하며, 이는 예측 모델의 성능에 악영향을 미칠 수 있다. 본 연구에서 사용된 코호트에서도 사례군이 전체의 약 6%로, 현저한 클래스 불균형이 관찰되었다. 불균형을 해소하기 위해, 대조군의 under-sampling 과 사례군의 over-sampling 기법을 적용하여 그 효과를 분석하였다.

Table 7 은 대조군 크기를 사례군의 1 배에서 4 배까지 조정한 결과를 나타낸다. 대조군 크기 증가에 따라 AUROC 는 점진적으로 상승하였으나, F1 점수는 감소하는 경향을 보였다. 그러나 이러한 성능 변화의 폭은 상대적으로 작았으며, 원 데이터셋과 조정된 데이터셋 간의 AUROC 차이는 통계적으로 유의미하지 않았다 (모든 비교에서 $P > 0.05$).

Table 8 은 사례군 크기를 원래의 2 배 및 3 배로 증강시킨 결과를 제시한다. 증강 방법으로는 각 환자의 첫 번째와 마지막 심전도를 포함하고, 중간 심전도를 무작위로 추출하는 샘플링 기반 기법을 사용하였다. 2 배 증강 시, AUROC (0.939 vs. 0.925, $P < 0.001$), AUPRC (0.575 vs. 0.532), NPV (0.984 vs. 0.981)에서 유의미한 성능 향상이 관찰되었다. 반면, F1 점수 (0.522 vs. 0.532)와 PPV (0.385 vs. 0.410)는 소폭 감소하였다. 3 배 증강의 경우, 전반적으로 원 데이터셋보다 낮은 성능을 나타냈다.

이러한 결과는 클래스 불균형 문제를 해결하기 위해 적용한 기법들이 원 데이터셋을 사용했을 때와 비교하여 성능 차이가 크지 않음을 보여준다. 이는 본 연구에서 제안한 모델이 클래스 불균형에 대해 견고성을 가지고 있으며, 불균형한 데이터 분포에도 불구하고 효과적으로 학습되었음을 시사한다.

Table 7. Sensitivity analysis according to different case:control ratios

Case: Control	Case samples	Control samples	AUROC	p-value	F1	AUPRC	PPV	NPV
1:1	1,783	1,782	0.916	0.438	0.872	0.916	0.866	0.847
1:2	1,783	3,566	0.919	0.519	0.801	0.860	0.737	0.916
1:3	1,783	534	0.928	0.799	0.776	0.818	0.715	0.933
1:4	1,783	7,132	0.936	0.180	0.758	0.824	0.718	0.937
default	1,783	30,233	0.925	-	0.532	0.572	0.41	0.981

Table 8. Sensitivity analysis according to different augmentation ratio

Augmentation ratio	Development set case samples	Development set control samples	AUROC	p-val	F1	AUPRC	PPV	NPV
no augmentation	1,280	23,512	0.925	-	0.532	0.572	0.410	0.981
x 2	2,560	23,512	0.939	<0.001	0.522	0.575	0.385	0.984
x 3	3,840	23,512	0.935	0.014	0.523	0.569	0.395	0.982

3.8 추가분석

본 섹션에서는 모델의 성능과 관련된 두 가지 추가 분석을 수행하였다. 첫째, 과거 심초음파 기록이 예측 성능에 미치는 영향을 조사하였고, 둘째, 데이터 양이 환자 임베딩에 미치는 영향을 평가하였다.

모델의 예측 성능이 과거 LVEF 감소 여부와 관련이 있는지 평가하기 위해 검증 세트를 두 그룹으로 나누어 분석하였다. 'Constant' 그룹은 과거와 최종 LVEF 감소 여부가 일치하는 환자들로, Control 은 지속적 정상, Case 는 지속적 감소인 환자를 나타낸다. 'Change' 그룹은 과거와 최종 기록 사이에 변화가 있는 환자들로, Control 은 감소에서 정상으로, Case 는 정상에서 감소로 변화한 경우를 의미한다 (Table 9). 분석 결과, Constant 그룹에서 모델의 성능이 현저히 높게 나타났다. 반면, 과거 기록이 마지막 기록과 불일치할 경우 모델의 예측 정확도가 감소하였다. 이는 모델이 과거 LVEF 상태를 반영하여 예측을 수행함을 시사한다. 그러나 이러한 결과는 동시에 모델이 LVEF 상태의 동적인 변화를 정확히 포착하는 데 있어 한계가 있음을 나타내며, 이 부분에 대한 추가적인 개선이 필요함을 시사한다.

연구의 프레임워크는 환자 별 임베딩을 생성할 때 고정된 max length 와 가중 합 방식을 사용한다. 이 과정에서 데이터가 없는 부분에는 0 의 가중치를 부여한다. 이로 인해 데이터 수가 많은 환자의 임베딩 값이 커질 수 있어, 사례군 환자들의 측정 기록이 대조군보다 많다는 점에서 모델이 임베딩 값의 크기만으로 예측을 수행할 수 있다는 우려가 제기되었다. 그러나 프레임워크는 가중 합 이후 정규화 과정을 거치므로 이러한 문제를 방지한다.

이를 검증하기 위해 데이터 수에 따른 임베딩 값의 차이를 분석하였다. Figure 10 은 환자별 심전도 기록 수의 중앙값인 4 를 기준으로, 4 개 이상인 그룹과 4 개 미만인 그룹의 임베딩 평균값 분포를 비교한 것이다. 분석 결과, 두 그룹 간 분포가 거의 일치하는 것으로 나타났다. 이는 모델이 임베딩 값의 크기가 아닌 실제 데이터에 내재된 정보를 바탕으로 예측을 수행함을 시사한다.

Table 9. Performance comparison between Constant and Change groups in LVEF status

Group	Case	Control	AUROC	F1	AUPRC	PPV	NPV
Constant	438	6,215	0.946	0.599	0.682	0.774	0.227
Change	65	506	0.628	0.261	0.195	0.187	1.000

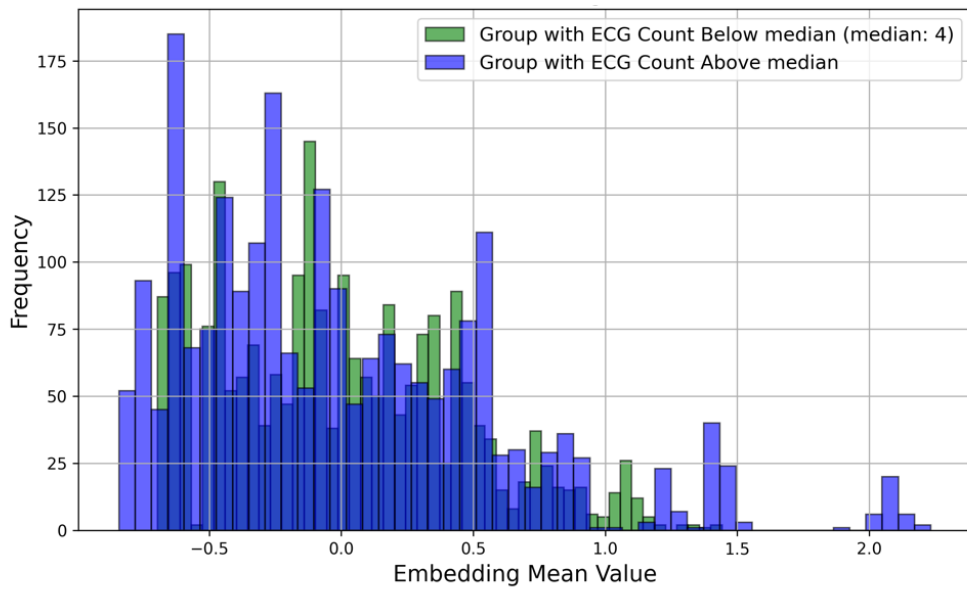


Figure 10. Comparison of embedding mean values based on ECG record count (Green) Group with ECG count below median (< 4) (Blue) Group with ECG count above or equal to median (≥ 4)
 Abbreviations: ECG, Electrocardiogram

제 4 장 고찰

본 연구는 종단 심전도 데이터를 활용한 딥러닝 기반 좌심실 박출률 감소 예측 모델을 개발하고 그 성능을 평가하였다. 가중 합 기반의 종단 심전도 분석 프레임워크를 고안하여 좌심실 박출률 감소 감지 성능을 개선하고자 하였으며, 다양한 실험을 통해 그 효과성을 검증하였다.

4.1 실험 결과 고찰

Baseline characteristics 분석을 통해 좌심실 박출률 감소 환자군과 대조군 간의 유의미한 임상적, 전기생리학적 차이를 확인하였다. 또한, 심전도 데이터 유사성 분석 결과는 환자 개인의 고유한 심전도 패턴이 시간에 따라 일관성을 유지함을 보여주어, 본 연구의 종단 분석 접근법의 타당성을 뒷받침하였다.

연구 결과, 제안된 종단 분석 방법(Longitudinal setting)이 단일 시점 분석 방법(Cross-sectional setting)보다 일관되게 우수한 성능을 보였다. 특히 ECGNet 모델을 사용한 종단 분석에서 가장 우수한 성능을 보였다.

기계학습 모델을 활용한 종단 분석 방법과 단일 시점 분석 방법 비교에서도 종단 분석 방법이 좋은 성능을 보였다. 하지만 기계학습 모델 중 제일 우수한 LightGB와 ECGNet을 비교하였을 때 ECGNet이 대부분의 지표에서 더 높은 성능을 달성하였다. 이는 심전도 신호의 복잡한 패턴을 학습하는 데 있어 딥러닝의 강점을 확인시켜 주며,

기존의 특징 추출 기반 접근법보다 원시 데이터를 직접 학습하는 딥러닝 방식이 더 효과적일 수 있음을 시사한다.

Explainable AI 기법을 활용한 모델 해석 가능성 분석은 본 연구의 임상적 적용 가능성을 높인다. Expected Gradients 와 Integrated Gradients 기법을 통해 모델의 예측에 중요하게 작용하는 시점과 심전도 신호의 특정 부분을 식별할 수 있었다. 분석 결과, 모델이 임상적으로 중요한 특징을 기반으로 학습하고 있음을 확인하였다. 이는 의료가 모델의 예측 결과를 이해하고 신뢰하는 데 도움을 줄 수 있으며, 나아가 환자 맞춤형 치료 계획 수립에 활용될 수 있는 가능성을 제시한다.

하이퍼파라미터 분석을 통해 모델 성능에 영향을 미치는 주요 요인들을 파악하였다. 적절한 감쇠율(decay rate) 설정, 깊은 weighted summation depth, 균일한 샘플링 기법이 성능 향상에 기여함을 확인하였다. 이러한 결과는 향후 유사한 종단 분석 모델 개발 시 참고할 수 있는 중요한 지침이 될 수 있다.

또한, 추론 시 환자의 과거 심전도 데이터를 추가할수록 성능이 점진적으로 향상되는 것을 관찰하였다. 이는 본 모델이 단순히 데이터의 양적 증가를 넘어서, 환자의 시간에 따른 심전도 변화 패턴을 효과적으로 학습하고 활용하고 있음을 시사한다. 이러한 결과는 종단 데이터 활용의 중요성을 재확인시켜 주며, 시간에 따른 질병 진행 과정의 정보가 모델 성능 개선에 크게 기여함을 의미한다.

4.2 한계점 및 발전 방향

본 연구의 한계점과 이를 개선하기 위한 향후 연구 방향은 다음과 같다.

가중치 함수 개선: 현재 모델은 시점 간 시간 차이(time delta)만을 고려하여 가중치를 부여하고 있다. 향후 연구에서는 마지막 심전도와의 유사도 값이나 방문 시점별 기계관독 특징 등을 반영하여 가중치 함수를 개선할 필요가 있다. 이를 통해 각 방문의 임상적 중요도를 더욱 정확히 반영할 수 있을 것이다.

다중 추상화 단계 활용: 현재 방문 별 심전도를 weighted summation 할 때 하나의 추상화 단계 데이터만을 사용하고 있다. 여러 추상화 단계의 데이터를 결합하여 사용하는 방법(예: weighted summation depth 가 1 인 임베딩과 3 인 임베딩을 결합)을 탐색하여 모델의 표현력을 향상시킬 수 있을 것이다.

최적 Max length 탐색: 현재 설정된 Max length 에 대한 최적화가 필요하다. 다양한 Max length 값에 따른 성능 변화를 분석하여 최적의 값을 도출해야 한다.

다중 모달리티 데이터 통합: 현재 모델은 심전도 데이터만을 사용하고 있다. 향후 연구에서는 실험실 검사 결과, 동반질환 정보 등 다른 모달리티의 데이터를 함께 활용하여 모델의 예측력을 향상시킬 필요가 있다.

다양한 태스크에 대한 적용: 제안된 프레임워크의 견고성(robustness)을 평가하기 위해 다양한 태스크에 적용해볼 필요가 있다. 예를 들어, 좌심실 박출량 값을 직접 예측하는 회귀 문제나 심부전 외 다른 심혈관 질환 예측 등에 적용해볼 수 있다.

다기관 연구를 통한 일반화 검증: 본 연구는 단일 기관의 데이터를 사용하였기 때문에 일반화 가능성에 제한이 있을 수 있다. 향후 다기관 연구를 통해 모델의 일반화 성능을 검증하고 개선할 필요가 있다.

클래스 불균형 문제 해결: 데이터셋의 클래스 불균형 문제가 존재하므로, 이를 해결하기 위해 논문에서 사용한 방법 외에 다양한 방법을 적용하여 모델의 성능을 더욱 개선할 수 있을 것이다.

이러한 한계점들을 극복하고 제안된 방향으로 연구를 발전시킨다면, 본 연구에서 개발한 종단 심전도 분석 모델의 임상적 유용성과 적용 가능성을 크게 향상시킬 수 있을 것으로 기대된다.

제 5 장 결론

본 연구는 딥러닝과 종단 심전도 데이터 분석을 통해 좌심실 박출률 감소를 예측하는 모델을 개발하였다. 심부전의 주요 원인인 좌심실 수축 기능 저하의 조기 진단과 지속적인 모니터링의 중요성을 인식하고, 기존 방법들의 한계를 극복하고자 하였다.

제안된 가중 합 기반 종단 심전도 분석 프레임워크는 다음과 같은 주요 성과를 달성하였다:

1. 종단 원시 심전도 데이터를 활용한 최초의 딥러닝 기반 좌심실 박출률 감소 예측 모델을 개발하였다.
2. 다중 방문 환자의 원본 심전도 신호를 직접 입력으로 사용하여 심전도에 내재된 다양한 정보를 최대한 활용하였다.
3. 지수 가중 함수를 통해 시간에 따른 데이터의 중요도를 효과적으로 반영하였다.
4. 종단 분석 방법이 단일 시점 분석 방법보다 일관되게 우수한 성능을 보였으며, 특히 ECGNet 모델을 사용한 종단 분석에서 최고의 성능을 달성하였다.
5. 심전도 검사 횟수 증가에 따른 성능 향상을 통해 종단 데이터 활용의 효과성을 입증하였다.
6. Explainable AI 기법을 통해 모델의 의사결정 과정에 대한 해석 가능성을 제공하여 임상적 신뢰도를 높였다.
7. 다양한 모델 아키텍처와 태스크에 적용 가능한 유연한 프레임워크를 제안하였다.

본 연구의 결과는 좌심실 박출률 감소의 조기 진단, 예후 예측, 맞춤형 치료 계획 수립 등 임상 실무에 직접적으로 적용될 수 있는 가능성을 제시한다. 특히, 시간에 따른 심전도 변화 패턴을 학습하고 활용함으로써, 환자 개개인의 특성을 고려한 정밀 의료 실현에 한 걸음 더 다가갈 수 있게 되었다.

그러나 본 연구에는 단일 기관 데이터 사용, 클래스 불균형 등의 한계점이 존재한다. 향후 연구에서는 다기관 데이터를 활용한 모델 검증, 클래스 불균형 문제 해소, 다중 모달리티 데이터 통합, 다양한 심혈관 질환으로의 적용 확장 등을 통해 모델의 일반화 가능성과 임상적 유용성을 더욱 높일 수 있을 것이다.

결론적으로, 본 연구에서 개발된 종단 심전도 분석 모델은 심부전 환자 관리의 질적 향상과 정밀 의료 실현에 실질적으로 기여할 수 있는 잠재력을 가지고 있다. 이는 심혈관 질환 관리에 있어 중요한 진전을 이루었음을 의미하며, 향후 다양한 후속 연구를 통해 더욱 발전된 형태로 임상 현장에 적용될 수 있을 것으로 기대된다.

참고 문헌

1. Shahim, B., et al., *Global public health burden of heart failure: an updated review*. Card Fail Rev, 2023. **9**.
2. Bozkurt, B., et al., *Heart failure epidemiology and outcomes statistics: a report of the Heart Failure Society of America*. J Card Fail, 2023. **29**(10): p. 1412.
3. Al-Mohammad, A., et al., *Diagnosis and management of adults with chronic heart failure: summary of updated NICE guidance*. BMJ, 2010. **341**.
4. McMurray, J.J., et al., *Left ventricular systolic dysfunction, heart failure, and the risk of stroke and systemic embolism in patients with atrial fibrillation: insights from the ARISTOTLE trial*. Circ Heart Fail, 2013. **6**(3): p. 451-460.
5. Vasan, R.S., et al., *Epidemiology of left ventricular systolic dysfunction and heart failure in the Framingham study: an echocardiographic study over 3 decades*. JACC Cardiovasc Imaging, 2018. **11**(1): p. 1-11.
6. Ito, S., et al., *Reduced left ventricular ejection fraction in patients with aortic stenosis*. J Am Coll Cardiol, 2018. **71**(12): p. 1313-1321.
7. Yancy, C.W., et al., *2017 ACC/AHA/HFSA focused update of the 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America*. Circulation, 2017. **136**(6): p. e137-e161.
8. Goel, S., et al., *Decline in left ventricular ejection fraction following anthracyclines predicts trastuzumab cardiotoxicity*. JACC Heart Fail, 2019. **7**(9): p. 795-804.

9. Wehner, G.J., et al., *Routinely reported ejection fraction and mortality in clinical practice: where does the nadir of risk lie?* Eur Heart J, 2020. **41**(12): p. 1249–1257.
10. Wang, T.J., et al., *Natural history of asymptomatic left ventricular systolic dysfunction in the community.* Circulation, 2003. **108**(8): p. 977–982.
11. McDonagh, T.A., K. McDonald, and A.S. Maisel, *Screening for asymptomatic left ventricular dysfunction using B-type natriuretic peptide.* Congest Heart Fail, 2008. **14**: p. 5–8.
12. Redfield, M.M., et al., *Burden of systolic and diastolic ventricular dysfunction in the community: appreciating the scope of the heart failure epidemic.* JAMA, 2003. **289**(2): p. 194–202.
13. Morgan, S., et al., *Prevalence and clinical characteristics of left ventricular dysfunction among elderly patients in general practice setting: cross sectional survey.* BMJ, 1999. **318**(7180): p. 368–372.
14. Real, J., E. Cowles, and A.S. Wierzbicki, *Chronic heart failure in adults: summary of updated NICE guidance.* BMJ, 2018. **362**.
15. Richards, M., et al., *Diagnostics: a major priority for the NHS.* Future Healthc J, 2022. **9**(2): p. 133.
16. Davie, A., et al., *Value of the electrocardiogram in identifying heart failure due to left ventricular systolic dysfunction.* BMJ, 1996. **312**(7025): p. 222–223.
17. Potter, E.L., et al., *Machine learning of ECG waveforms to improve selection for testing for asymptomatic left ventricular dysfunction.* JACC Cardiovasc Imaging, 2021. **14**(10): p. 1904–1915.
18. Kwon, J.-m., et al., *Development and validation of deep-learning algorithm for electrocardiography-based heart failure identification.* Korean Circ J, 2019. **49**(7): p. 629.

19. Choi, J., et al., *Deep learning of ECG waveforms for diagnosis of heart failure with a reduced left ventricular ejection fraction*. Sci Rep, 2022. **12**(1): p. 14235.
20. Attia, Z.I., et al., *Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram*. Nat Med, 2019. **25**(1): p. 70–74.
21. Adedinsowo, D., et al., *Artificial intelligence-enabled ECG algorithm to identify patients with left ventricular systolic dysfunction presenting to the emergency department with dyspnea*. Circ Arrhythm Electrophysiol, 2020. **13**(8): p. e008437.
22. Mondo, C., et al., *External validation of an electrocardiography artificial intelligence-generated algorithm to detect left ventricular systolic function in a general cardiac clinic in Uganda*. Eur Heart J, 2020. **41**(Supplement_2): p. ehaa946. 1013.
23. Cho, J., et al., *Artificial intelligence algorithm for screening heart failure with reduced ejection fraction using electrocardiography*. ASAIO J, 2021. **67**(3): p. 314–321.
24. Kashou, A.H., et al. *Artificial intelligence-augmented electrocardiogram detection of left ventricular systolic dysfunction in the general population*. in *Mayo Clin Proc*. 2021. Elsevier.
25. Attia, Z.I., et al. *Artificial intelligence ECG to detect left ventricular dysfunction in COVID-19: a case series*. in *Mayo Clin Proc*. 2020. Elsevier.
26. Feng, S., et al., *Machine learning-aided detection of heart failure ($LVEF \leq 49\%$) by using ballistocardiography and respiratory effort signals*. Front Physiol, 2023. **13**: p. 1068824.
27. Vaid, A., et al., *Using deep-learning algorithms to simultaneously identify right and left ventricular dysfunction from the*

- electrocardiogram*. JACC Cardiovasc Imaging, 2022. **15**(3): p. 395–410.
28. Taylor, C., et al., *The potential role of NT-proBNP in screening for and predicting prognosis in heart failure: a survival analysis*. BMJ Open, 2014. **4**(4): p. e004675.
 29. Lou, Y.-S., et al., *Development and validation of a dynamic deep learning algorithm using electrocardiogram to predict dyskalaemias in patients with multiple visits*. Eur Heart J Digit Health, 2023. **4**(1): p. 22–32.
 30. Pham, H.N., et al., *Dynamic electrocardiogram changes are a novel risk marker for sudden cardiac death*. Eur Heart J, 2024. **45**(10): p. 809–819.
 31. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*. 2016.
 32. Dosovitskiy, A., et al., *An image is worth 16x16 words: Transformers for image recognition at scale*, in *International Conference on Learning Representations (ICLR)*. 2021.
 33. Yoo, J., et al., *Residual one-dimensional convolutional neural network for neuromuscular disorder classification from needle electromyography signals with explainability*. Comput Methods Programs Biomed, 2022. **226**: p. 107079.
 34. Fluss, R., D. Faraggi, and B. Reiser, *Estimation of the Youden Index and its associated cutoff point*. Biom J, 2005. **47**(4): p. 458–472.
 35. Cramer, J.S., *The origins of logistic regression*. 2002.
 36. Fisher, R.A., *The use of multiple measurements in taxonomic problems*. Annals of eugenics, 1936. **7**(2): p. 179–188.

- 37. Ke, G., et al. *LightGBM: a highly efficient gradient boosting decision tree*. in *Advances in Neural Information Processing Systems (NIPS) Workshop*. 2017.
- 38. Geurts, P., D. Ernst, and L. Wehenkel, *Extremely randomized trees*. Mach Learn, 2006. **63**: p. 3-42.
- 39. Erion, G., et al., *Improving performance of deep learning models with axiomatic attribution priors and expected gradients*. Nat Mach Intell, 2021. **3**(7): p. 620-631.
- 40. Sundararajan, M., A. Taly, and Q. Yan. *Axiomatic attribution for deep networks*. in *International conference on machine learning (ICML)*. 2017. PMLR.

Abstract

Longitudinal ECG Analysis

with Deep Learning for Improved Detection of Reduced Left Ventricular Ejection Fraction

Soomin Chung

Interdisciplinary Program in Bioengineering

The Graduate School

Seoul National University

Background: Heart failure is a serious global health problem, with left ventricular systolic dysfunction (LVSD) as a major cause. Existing echocardiogram is limited by operator-dependency and subjective interpretation. Previous electrocardiogram (ECG)-based LVSD diagnosis studies using single-point data fail to consider temporal changes and individual characteristics.

Methods: We developed a deep learning model to predict decreased left ventricular ejection fraction (LVEF) using longitudinal ECG data. The proposed weighted sum-based framework uses raw ECG signals from multiple patient visits as input and is adaptable to various model architectures. An exponential weighting function assigns weights

based on time intervals between data points, emphasizing recent data while utilizing historical information.

Results: The proposed longitudinal method consistently outperformed cross-sectional approaches, with the ECGNet model achieving the highest performance. Model performance improved with increasing ECG examination frequency, demonstrating the effectiveness of longitudinal data utilization. Comparisons with machine learning models confirmed the superiority of our deep learning approach. Model interpretability analysis using Expected Gradients techniques provided clinical insights into the model's decision-making process.

Conclusion: This study developed an model for predicting decreased LVEF through deep learning and longitudinal ECG data analysis. The model shows promise for early diagnosis, prognosis prediction, and personalized treatment planning in heart failure patients. However, limitations include the use of single-institution data and class imbalance. Future research should focus on multi-center studies, multi-modal data integration, and application to various cardiovascular diseases to enhance the model's generalizability and clinical utility.

Keywords: Heart failure, Low left ventricular ejection fraction, Electrocardiogram, Deep learning, Longitudinal data analysis

Student Number: 2022-27947