

심전도를 이용한 딥러닝 기반 좌심실 비대 분류

류지승¹, 강현영¹, 이예린¹, 서경덕¹, 이솔암^{2,3}, 박영준^{4*}, 양세정^{1*}

연세대학교 의공학과¹

연세대학교 원주의과대학 피부과학 교실²

연세대학교 원주의과대학 예방의학 교실³

연세대학교 원주의과대학 내과학 교실⁴

Left Ventricular Hypertrophy Classification based on Deep Learning using Electrocardiography

Jiseung Ryu¹, Hyunyoung Kang¹, Yerin Lee¹, Kyungdeok Seo¹, Solam Lee^{2,3}, Youngjun Park^{4*}, Sejung Yang^{1*}

Department of Biomedical Engineering, Yonsei University, Wonju, Republic of Korea¹

Department of Dermatology and Institute of Hair and Cosmetic Medicine, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea²

Department of Preventive Medicine, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea³

Division of Cardiology, Department of Internal Medicine, Wonju Severance Christian Hospital, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea⁴

fbwkdrns77@yonsei.ac.kr¹, sonya23@yonsei.ac.kr¹, yerin001@hotmail.com¹, rud395@yonsei.com¹, solam@yonsei.ac.kr^{2,3}, pyj@yonsei.ac.kr^{4*}, syang@yonsei.ac.kr^{1*}

Abstract

Left Ventricular Hypertrophy(LVH) is defined as a type of response of the myocardium to abnormal heart load. LVH is considered a risk factor for cardiovascular disease and is closely related to the onset and prognosis of the disease. Therefore, it is very clinically important to accurately diagnose LVH, an early stage of heart change. Electrocardiography(ECG) is the easiest, non-invasive, and economical diagnostic method among LVH diagnosis methods. ECG is widely used as a screening test for LVH, but the coincidence rate of actual left ventricular hypertrophy and diagnosis findings is low. Recently, algorithms and models using big data and deep learning have been in the spotlight. In this study, we developed attention based deep learning model that can diagnose LVH. Here, we develop a deep learning model that enables end-to-end learning with only a given ECG and can classify LVH by considering demographic features and ECG features.

1. 연구 배경

좌심실 비대(Left Ventricular Hypertrophy, LVH)는 비정상적인 심부하에 따른 심근의 한 반응형태이다. 좌심실 비대가 동반된 경우 심혈관계 위험이 5배에서 10배가량 증가하는 것으로 알려져 있으며, 심근경색 환자에서도 유사한 결과를 보이는 것으로 알려져 있다. 좌심실 비대는 심부전 및 부정맥 등 심혈관 질환의 위험 인자로 여겨지며, 발병 및 예후와 밀접한 관계가 있다. 그러므로, 심장변화의 초기 단계인 좌심실 비대를 정확히 진단하는 것은 고혈압 환자의 치료로 이어질 수 있기에 임상적으로 매우 중요하다.[1]

현재 좌심실 비대를 진단하는 방법으로는 단순 흉부 방사선 검사(X-ray), 심전도 검사(Electrocardiography, ECG), 심장초음파 검사(Echocardiography), 좌심실 조영술(Contrast ventriculography), 자기 공명 영상(Magnetic Resonance Imaging) 등이 있다. 이 중 가장 손쉽고 비침습적이며 경제적인 진단 방법인 심전도 검사가 좌심실 비대의 선별 검사(Screening test)로 널리 사용되고 있다.

심전도 검사에서 좌심실 비대기준의 척도는 주로 QRS 전압을 이용한다. 따라서, Sokolov-Lyon voltage standard, Cornell voltage standard 등 여러 기준표가 제안되었지만,[2] QRS 전압의 크기에만 근거하였기 때문에 실제 좌심실 비대 여부와 진단 소견의 일치율은 낮았다. 따라서, 머신 러닝(Machine Learning, ML)을 이용한 좌심실 비대 분류가 연구되었으나 마찬가지로 낮은 분류 결과를 보였다.[3-4]

최근에는 축적된 빅데이터와 딥러닝(Deep Learning, DL)을 이용한 알고리즘 및 모델이 주목받고 있다. 주어진 정보를 스스로 학습하고 분석하여 분류에 필요한 특징을 자동

추출하여 기존의 알고리즘에 의존한 머신 러닝 기법의 단점을 보완하고, 의료 분야에서의 광범위한 분류 및 식별 작업에서도 최첨단 성능을 달성했다.[5-6] 따라서, 본 연구에서는 연세대학교 원주 세브란스 병원에서 2010년에서 2020년까지 축적된 심전도 빅데이터를 사용하여 좌심실 비대를 분류하기 위한 딥러닝 모델을 개발하였다. 여기서 우리는 주어진 심전도 데이터만으로 End-to-End 학습이 가능한 모델을 설계하였으며, 다층 퍼셉트론(Multi-Layer Perceptron, MLP)를 통한 환자의 인구 통계학적 요소들과 심전도의 특징들에 대한 고려가 가능하도록 하였다. 또한, 이미지 분류 분야에서 활발히 사용되고 있는 Self-attention 기법을 적용하여, 모델이 분류에 필요한 특징에 더욱 집중할 수 있도록 설계하였다.

2. 연구 방법

2-1) 전처리(Preprocessing)

본 연구에는 2010년 10월에서 2020년 2월까지 연세대학교 원주 세브란스 기독교 병원에서 획득된 심전도 데이터가 사용되었다. 기록된 심전도는 8-Lead (I, II, V1, V2, V3, V4, V5, V6)이며, 4-Lead (III, aVF, aVR, aVL)를 벡터 계산을 통해 추가하여 12-Lead를 구성하였다. 원본 데이터는 샘플링 레이트 500Hz로 10초 동안 기록된 데이터로 총 5,000 샘플로 이루어져 있다. 전후 1초 내에는 심전도 측정 시 생기는 노이즈가 다소 많이 포함되어 있기에, 이를 제외한 총 8초의 데이터가 실험에 사용되었다. Normalization이나 Noise Reduction 등 신호처리를 가한 데이터로 실험을 진행하였을 때, 딥러닝 모델의 결과간

유의미한 차이를 보여주지 않았다. 따라서, 본 연구에서는 신호처리를 가하지 않은 원본 심전도가 사용되었다.

2-2) 라벨링(Labeling)

데이터의 라벨을 구체적으로 정의하기 위해서, 심장초음파 검사를 통해 측정된 좌심실 질량을 체표면적으로 나눈 'LV mass index/BSA by 2D method'가 사용되었다. 라벨은 심장초음파 수치와 남녀의 심장차이를 고려한 기준표를 통해 4개의 세부클래스로 분류되었으며, 선별 검사로서 의미가 있는 Mildly abnormal과 Moderately abnormal 사이를 Cut-off 기준으로 정하였다. 최종적으로, 남성의 경우 $132g/m^2$, 여성의 경우 $109g/m^2$ 보다 작으면 정상, 크면 좌심실 비대라 정의하여 각각 0과 1로 라벨링 되었다.

2-3) 데이터셋(Data set)

심장초음파 검사를 통해 얻은 라벨에 대해서 한 명의 환자가 여러 번 검사를 받았다면, Initial 라벨을 사용하여, 중복되는 인원이 없도록 처리를 진행하였다. 이후, 심장초음파 측정일 전후 31일 이내로 심전도 검사를 하였다면, 기록된 심전도와 쌍을 이루어 데이터셋을 구성할 수 있도록 하였다. 마지막으로, 각각의 검사 시, 비정상적으로 측정된 심전도 데이터와 심장 초음파 라벨을 제외하였다. 최종적으로 37,062명의 환자가 데이터셋으로 사용되었다. 데이터셋은 4:3:3의 비율로 분할 되었으며, 중복된 인원은 없었다. Train set과 Validation set은 딥러닝 모델의 파라미터를 조정하기 위해 사용되었으며, Test set은 딥러닝 모델의 최종 성능을 평가하는데 사용되었다.

	Train	Validation	Test
Ratio	4	3	3
Total	14,822	11,119	11,121
Normal	12,698	9,525	9,526
LVH	2,124	1,594	1,595

표 1. 데이터셋 구성

2-4) 모델 설명(Model description)

딥러닝 모델은 좌심실 비대와 관련된 선행 연구와 이미지 분류 분야에서 사용되는 선행 연구들을 참고하여 설계되었다. 딥러닝 모델은 Feature Extract 단은 CNN(Convolutional Neural Networks)과 다층 퍼셉트론으로 구성되었다. 우선 심전도 데이터는 Convolution Block와 Attention Block을 반복적으로 통과하며, 분류에 필요한 Features를 추출한다. Convolution Block은 일반적인 CNN에서 사용되는 Convolution, Batch Normalization, Max Pooling layer로 구성되어 있으며, Local Features를 찾는 역할을 수행하였다. Attention Block은 Channel Attention과 Spatial Attention을 반복적으로 수행하여, 중요도가 높은 부분을 강조하여 효과적으로 학습이 이루어질 수 있도록 도움을 주는 역할을 수행하였다. 이후 2개의 Demographic Features와 9개의 심전도 Features는 5개의 Hidden layer와 8개의 출력 노드로 이루어진 다층 퍼셉트론을 통해 정제되었다. CNN의 Local Features와 다층 퍼셉트론의 Features는 합쳐져서 Fully Connected layer로 전달되었다. 최종적으로 딥러닝 모델은 환자가 좌심실 비대에 해당될

확률을 출력하였다. 출력된 확률이 50%보다 크다면 모델의 예측 값은 1로 정의되었다.

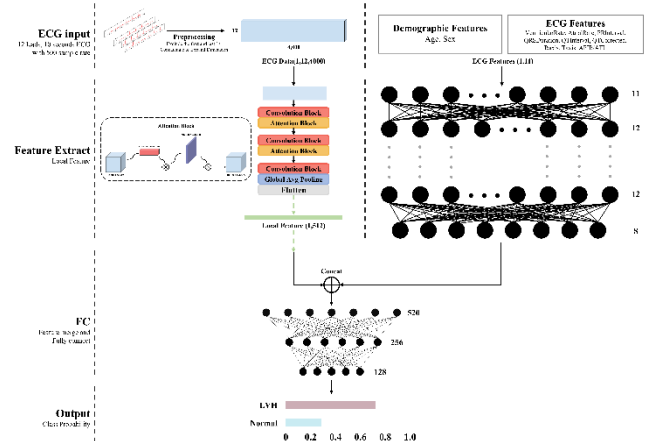


그림 1. 딥러닝 모델

2-5) 하이퍼 파라미터(Hyper parameters)

손실 함수로는 이진 분류에서 사용되는 Binary Cross Entropy(BCE) loss가 사용되었다. Train set 내 좌심실 비대 환자의 수가 정상 환자의 수에 비해 상대적으로 매우 적기 때문에, 학습과정에서 이를 보완해주고자 상대 비율을 계산하여 손실 함수에 적용하였다. 초기 학습율의 범위는 0.01부터 0.0001까지, 초기 배치 사이즈(Batch size)의 범위는 64부터 512까지였다. 하지만 최종 성능간 유의미한 차이는 없었다. 따라서, 선행 연구의 참고와 Grid searches를 통해, 학습률은 0.005, 배치 사이즈는 128로 조정되었다. 딥러닝 모델은 Adam optimizer를 활용하여 100 epochs 동안 학습되었다. 학습 과정 동안 Validation set의 성능이 가장 높을 때, 최적의 모델로 설정되어 저장되었으며, 이후 저장된 모델을 불러와 Test set을 통해 딥러닝 모델의 최종 성능을 평가하였다. 모델 학습을 위한 프레임워크는 Pytorch와 NVIDIA GeForce RTX 2080 Ti가 사용되었다.

3. 연구 결과

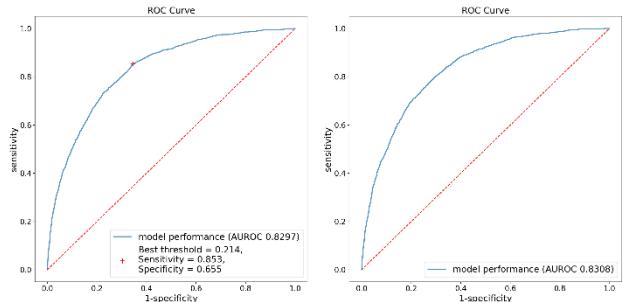


그림 2. Validation set의 AUROC 와 Test set의 AUROC

3-1) 평가 지표(Evaluation metrics)

평가 지표로는 질환을 양성으로 진단하는 비율인 민감도 (Sensitivity), 음성으로 진단하는 비율인 특이도(Specificity), 양성으로 진단 시, 실제 양성일 확률인 양성 예측치(Positive Predictive Value, PPV), 음성으로 진단 시, 실제 음성일 확률인 음성 예측치(Negative Predictive Value, NPV), 마지막으로 민감도와 특이도를 선택하는 판단 기준들을 그린 ROC curve(Receiver Operating Characteristic curve) 내 면적을 의미하는 AUROC(Area Under the Receiver Operating

characteristics Curve)가 사용되었다. 본 연구에서는 AUROC를 주요 지표로 사용하였다.

3-2) 모델 성능(Model performance)

그림3는 딥러닝 모델의 AUROC를 나타낸다. 딥러닝 모델은 Validation set에 대해서 AUROC 0.8297, Test set에 대해서 AUROC 0.8308을 보였다. 다른 성능 지표에 대해서는 민감도 55.36%, 특이도 87.70%, 양성 예측치 42.98%, 음성 예측치 92.14%을 보였다.

3-3) 결론(Conclusion)

결과적으로, 우리가 설계한 딥러닝 모델의 학습에 사용된 데이터셋은 상대적으로 클래스간 불균형이 심하였음에도 낮지 않은 AUROC를 보여주었다. 또한, 높은 특이도와 음성 예측치를 보여주며, 선별검사에도 충분히 사용될 수 있는 성능을 보여주었다. 정리하면, 우리는 본 연구를 통해 심전도 빅데이터와 딥러닝 모델을 사용하여, 좌심실 비대를 가진 환자와 정상 환자를 분류할 수 있으며, 이를 기반으로 좌심실 비대를 진단할 수 있음을 보여주었다.

4.참고 문헌

- [1]SungHee Yang, JinSoo Lee, Changsoo Kim, “The Accuracy of Echocardiography and ECG in the Left Ventricular Hypertrophy”, The Korea Contents Association, Vol.16, Issue 2, Pages.666- 672, 2016.
- [2]Bacharova L, Estes EH. Left Ventricular Hypertrophy by the Surface ECG. J Electrocardiol. 2017 Nov-Dec;50(6):906-908. doi: 10.1016/j.jelectrocard.2017.06.006. Epub 2017 Jun 7. PMID: 28651797.
- [3]Jothiramalingam, R., Jude, A., Patan, R. et al. Machine learning-based left ventricular hypertrophy detection using multi-lead ECG signal. Neural Comput & Applic 33, 4445–4455 (2021). <https://doi.org/10.1007/s00521-020-05238-2>
- [4]Fernando De la GS, Maria Elena RI, Elias Abraham RD, et al., “Improvement of electrocardiographic diagnostic accuracy of left ventricular hypertrophy using a Machine Learning approach”, PLoS One, 2020 May 13;15(5): e0232657.
- [5]Z. Ebrahimi, M. Loni and M. Daneshtalab et al “A review on deep learning methods for ECG arrhythmia classification” Expert Systems with Applications: X Vol 7 , 2020.
- [6]JoonMyoung Kwon, KiHyun Jeon, HyueMee Kim, MinJeong Kim, et al., “Comparing the performance of artificial intelligence and conventional diagnosis criteria for detecting left ventricular hypertrophy using electrocardiography”, EP Europace, Vol.22, issue 3, Pages 412-419, 2020.