

양방향 LSTM을 활용한 질병 분류 모델

김현진¹, 김지은¹, 박현주¹, 김대환¹, 곽근창², 염찬욱³¹조선대학교 정보통신공학부 학부생²조선대학교 전자공학부 교수³조선대학교 AI융합대학사업단 연구교수

guswlsdl7771@chosun.kr, rlawldms1277@chosun.kr, qkrguswn1114@chosun.kr, kdha09@chosun.kr

kwak@chosun.ac.kr, walt18@chosun.ac.kr

Disease Classification Model Using Bidirectional LSTM

Hyun-Jin Kim¹, Ji-Eun Kim¹, Hyun-Ju Park¹, Dae-Hwan Kim¹,Keun-Chang Kwak², Chan-Uk Yeom³¹Undergraduate student in Information and Communication Engineering, Chosun University²Professor of Electronic Engineering, Chosun University³Research Professor of Division of AI Convergence College, Chosun University

요 약

본 연구는 ECG(Electrocardiogram)를 활용한 질병 분류 기술 개발을 목표로 1D CNN을 설계하여 ECG 데이터를 분석하였다. MIT-BIH Arrhythmia Database를 사용해 전처리 후 1D CNN 모델로 시계열 데이터에서 질병 분류에 유효한 특징을 추출하였다. 훈련 데이터에서는 약 81.91%의 정확도를 기록했으나, 검증 데이터에서는 57.5%로 성능 편차가 확인되었다. 이는 데이터 불균형의 영향으로 보이며, ECG가 비침습적이면서도 보안성이 높은 생체 신호로서 질병 분류에 유망하다는 것을 시사한다. 향후 연구에서는 다양한 데이터셋과 모델 고도화를 통해 성능을 개선할 계획이다.

1. 서론

현대 사회에서 질병 조기 진단과 예방은 매우 중요한 이슈로 부각되고 있다. 특히 심전도(ECG)는 심장의 전기적 활동을 측정하는 비침습적이고 신뢰성 높은 생체 신호로, 질병의 진단과 분류에 유용한 도구로 주목받고 있다[1]. 심전도는 개인의 심장 구조와 기능을 반영한 고유한 신호 패턴을 제공하며, 이를 통해 다양한 심혈관 질환의 분류가 가능하다[2]. 최근 심전도를 활용한 질병 분류 연구는 딥러닝 기술의 발전과 함께 크게 성장하고 있으며, 특히 CNN과 RNN을 결합한 모델이 높은 성능을 보이고 있다[3]. 본 논문에서는 ECG 데이터를 기반으로 한 질병 분류 모델을 제안하며, 1D CNN과 Bidirectional LSTM을 결합하여 심전도 신호에서 중요한 특징을 추출하고, 이를 통해 질병을 분류하는 방안을 제시한다. 기존 연구에서 지적된 데이터 불균형 문제를 해결하기 위해 정규화와 차원 조정 과정을 거쳤으며, 다양

한 성능 평가 지표를 활용하여 모델의 성능을 분석하였다[4]. 본 연구는 향후 심전도 기반 질병 진단 기술의 정확성을 높이는 데 기여할 것으로 기대된다[5].

2. Bidirectional LSTM model

본 연구에서는 두 가지 ECG 데이터셋을 사용한다. 첫 번째 데이터셋은 140개의 데이터 포인트와 이진 레이블(정상/비정상)을 포함한다. 두 번째 데이터셋은 188개의 고정 차원을 가진 심전도 신호로, 다양한 심장 상태를 분류한다.

- **정규화**: 모든 ECG 신호를 0과 1 사이로 정규화하여 일관성을 확보하였다.
- **차원 조정**: 모든 샘플의 길이를 동일하게 맞추어 모델 입력 형식에 부합하도록 하였다.
- **데이터 분할**: 전체 데이터셋을 80% 훈련셋, 20% 테스트셋으로 분할하였다.

본 연구에서는 Bidirectional LSTM 모델을 사용하여 시계열 ECG 데이터를 분석한다. 모델은 두 개의 BLSTM 레이어와 시그모이드 활성화 함수를 가진 출력 레이어로 구성되어 있다.

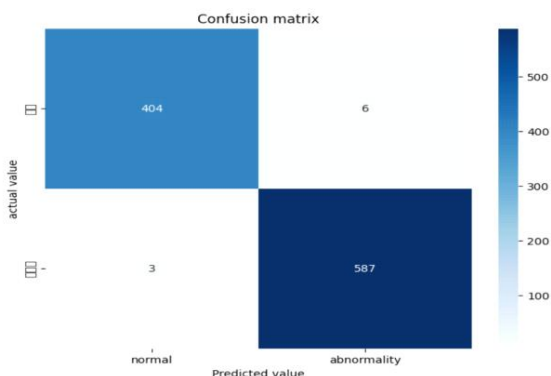
3. 결과 및 분석

본 연구에서는 ECG 데이터를 활용하여 질병 분류의 가능성을 평가하기 위한 실험을 진행하였다. 실험 결과는 다음과 같다.

- **정확도:** 모델의 훈련 정확도는 0.8191, 검증 정확도는 0.5750으로 나타났다.

- **손실:** 손실 함수 값은 초기 1.0007에서 최종적으로 0.4877로 소하며, 모델의 학습이 성공적으로 이루어졌음을 보여준다. 평가 지표 모델의 성능을 평가하기 위해 민감도, 특이도 등의 지표를 분석하였으며, 혼동 행렬을 통해 확인한 바와 같이, 특정 클래스(N, S, V 등)는 높은 정확도를 보였으나, 일부 클래스는 성능이 낮아 향후 추가적인 데이터 수집과 모델 개선이 필요함을 확인하였다. 비교를 위해 MIT-BIH Arrhythmia Dataset을 사용한 경우, 테스트 정확도가 0.4928로 나타났다. 반면, Kaggle의 ECG Dataset을 활용했을 때는 테스트 정확도가 0.9890으로 매우 높은 성능을 보인다. 이는 (그림 1)과 같이 모델이 데이터를 정상으로 정확하게 예측한 사례 수가 404개, 정상으로 오인한 사례는 6개로 높은 정확도를 기록한다.

(그림 1) ECG Dataset의 혼동 행렬



4. 결론 및 향후 연구 방향

본 연구는 심전도(ECG)를 활용하여 다양한 심혈관 질환을 효과적으로 분류할 수 있는 가능성을 MIT-BIH Arrhythmia Database를 통해 평가하였다. 실험 결과, 훈련 정확도는 약 80%로 나타났으나, 검증 정확도는 약 50%에 그쳐 일부 질병 클

래스에서 낮은 분류 성능이 확인되었다. 손실 함수 값은 1.0007에서 0.4877로 감소하며 모델의 학습이 진행되었음을 보여준다. 이러한 결과는 심전도가 특정 질병의 생리적 특성을 잘 반영할 수 있음을 시사한다.

향후 연구에서는 다양한 환경에서의 데이터 수집 및 최신 딥러닝 기법 적용을 통해 성능을 개선하고, 특히 낮은 성능을 보였던 클래스에 대한 집중 연구를 통해 심전도 기반 질병 분류 기술의 신뢰성과 정확도를 더욱 향상시킬 계획이다.

Acknowledgement

이 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 지원을 받아 수행된 AI 헬스케어 융합 대학 사업 연구임.

참고문헌

- [1] Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., & Ng, A. Y., "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Medicine*, vol. 25, no. 1, pp. 65-69, 2019.
- [2] Acharya, U. R., Fujita, H., Lih, O. S., Hagiwara, Y., Tan, J. H., & Adam, M., "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network," *Information Sciences*, vol. 405, pp. 81-90, 2017.
- [3] Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., & Ng, A. Y., "Cardiologist-level arrhythmia detection with convolutional neural networks," *arXiv preprint arXiv:1707.01836*, 2017.
- [4] Zubair, M., Kim, J., & Yoon, C., "An automated ECG beat classification system using convolutional neural networks," *Journal of Mechanics in Medicine and Biology*, vol. 16, no. 1, pp. 1650013, 2016.
- [5] Clifford, G. D., Liu, C., Moody, B., Lehman, L. H., Silva, I., Li, Q., & Mark, R. G., "AF classification from a short single lead ECG recording: The PhysioNet/Computing in Cardiology Challenge 2017," *Computing in Cardiology (CinC)*, Rennes, France, 2017, pp. 1-4.