

심장질환 환자 ECG 데이터 분석을 위한

딥러닝 기법 설계 및 경량화 모델 구축



풀악셀

착수보고서

신다윗(201824510)

안주현(201824520)

이재현(201824554)

## 목차

1. 졸업과제의 목표
2. 대상 문제
3. 요구조건 분석서
4. 현실적 제약 사항 분석 결과 및 대책
5. 설계 문서
6. 추진 체계 및 일정
7. 구성원 역할 분담

## 1. 졸업과제의 목표

심장질환 환자 ECG 데이터 분석을 위한 딥러닝 기법 설계 및 경량화 모델 구축한다.

## 2. 대상 문제

심장질환 환자의 심장 박동 패턴을 분석하고 비정상적인 패턴을 감지하는 딥러닝 모델을 구축하는 것이다. 이러한 모델은 심장질환 환자의 심장 박동 패턴을 실시간으로 모니터링하는 장치에 설치되어 심장질환 환자를 도울 수 있다. 예를 들어 부정맥, 협심증, 심근경색 등의 질환을 앓고 있거나 고지혈증, 고혈압, 당뇨병 등 심혈관 건강과 관련된 질환을 앓고 있는 환자의 심장 건강 상태를 확인할 수 있다.

이러한 기능을 구현하기 위해서는 심장박동 패턴의 특성을 식별하고 분류할 수 있는 모델이 필요하다. CNN 모델을 이용하여 심전도 데이터를 분석하고 분류할 수 있는 딥러닝 모델을 구축하고 경량화 하는 것이 목표이다.

- 심전도(electrocardiogram, ECG)

정해진 시간에 심장의 전기적 활동을 해석하는 것이다. 심전도는 피부에 부착된 전극과 신체 외부의 장비에 의해 기록된다. 기록은 이렇게 몸에 침투하지 않는 장비로 만들어지며 이 과정에 대한 용어를 심전도라고 한다. 심전도 기록은 심장의 전기적 활동에 대한 기록이다.

심전도는 심장박동의 비율과 일정함을 측정하는데 사용할 뿐만 아니라, 심장의 크기와 위치, 심장의 어떠한 손상이 있는지의 여부, 그리고 심박조율기와 같이 심장을 조절하는 장치나 약과 같은 효과를 보기 위해 사용된다.

## 3. 요구조건 분석서

- ECG 데이터를 기반으로 하는 예측 모델 생성한다.
- 모델의 경량화  
하드웨어 리소스가 제한된 갤럭시 워치 및 임베디드 시스템(라즈베리 파이)과 같은 환경에서 동작할 수 있도록 가볍고 효율적인 모델 개발한다.

## 4. 현실적 제약 사항 분석 결과 및 대책

- 제약사항 및 대책

부산대학교병원 환자의 **EHR** 데이터를 사용한 딥러닝 기법 설계가 처음 목표였지만 부산대학교병원 외부에서 환자 데이터를 사용하는 것이 불가능하다. **EHR** 데이터 사용을 하기 위해서 공간적, 시간적 제약이 있어서 비교적 접근이 용이한 **ECG** 데이터를 사용하는 것으로 변경하였다.

- 전자건강기록(**Electronic Health Records, EHR**)

디지털 형태로 체계적으로 수집되어 전자적으로 저장된 환자 및 인구의 건강정보이다.

이러한 기록은 서로 다른 헬스케어 환경에서 공유될 수 있다.

기록들은 네트워크로 연결된, 엔터프라이즈 규모의 정보시스템 및 다른 정보 네트워크와 교환을 통해 공유된다.

**EHR**은 인구통계학, 병력, 약물복용 및 알레르기. 예방접종 상태, 검사실 검사결과, 영상의학 이미지, 생체징후, 나이와 성별 같은 개인적인 통계, 그리고 청구정보를 포함하는 일정 범위의 데이터를 포함할 수 있다.

## 5. 설계 문서

- 개발환경

Google Colab

- 개발언어

Python

- 개발도구

Python 라이브러리(TensorFlow, Keras, Scikit-learn)

Google Colab(안정적인 GPU 환경 제공)

- 실행환경

Windows

## 6. 추진 체계 및 일정

- 모델 구축

- CNN(Convolutional neural network) 구조를 갖는 딥러닝 모델  
이미지 처리 작업(이미지 분류 및 인식, 이미지 분할)에 적합하다.

- ☐ CNN 모델에 [Attention Mechanism](#)을 적용해보기

- ☐ [Attention Mechanism](#) 적용한 CNN 모델의 고도화시키기

- 모델 검증

교차 검증의 목표는 과적합 또는 선택 편향 과 같은 문제를 표시하기 위해 추정  
사용되지 않은 새로운 데이터를 예측하는 모델의 능력을 테스트한다.

모델이 독립적인 데이터 세트(예: 실제 문제에서 알 수 없는 데이터 세트)로 일반화되는  
방법에 대한 통찰력을 제공한다.

- K-Folds cross-validator 방식  
여러 반복에 대한 결과를 평균화하여 모델 성능을 보다 안정적으로 추정할  
수 있다.
- Stratified K-Folds cross-validator 방식  
각 fold에 데이터가 편향되지 않게 데이터의 분포를 균일하게 만들고  
학습시킬 수 있다.

- 모델 평가

분류 모델은 일반적으로 모델 예측의 정확도, 정밀도, 재현율 및 F1 점수를 측정하는  
다양한 메트릭을 사용하여 평가된다.

- accuracy
- sensitivity
- specificity
- precision
- F1-score
- ROC(Receiver operating characteristics)
- PR곡선

- 모델 경량화

- **Network Pruning**  
Pruning은 Model의 weight들 중 중요도가 낮은 weight의 연결을 제거하여  
모델의 파라미터를 줄이는 방법이다.
- **Knowledge Distillation**  
Teacher Network로부터 증류한 지식을 Student Network로 transfer하는  
일련의 과정이 있다.
- **Network Quantization**

weights, biases, and activations 값의 정확도를 낮춰 메모리 사용을 줄이는 방법이다.

개발 일정	상태	마감일
프로젝트 착수보고서 작성	진행 중 ▾	2023년 5월 22일
딥러닝 모델 구축 및 성능 테스트	시작되지 ... ▾	2023년 7월 28일
중간 보고서 작성	시작되지 ... ▾	2023년 7월 31일
딥러닝 모델 경량화	시작되지 ... ▾	2023년 9월 28일
최종 보고서 작성	시작되지 ... ▾	2023년 9월 30일
졸업과제 발표	시작되지 ... ▾	2023년 10월 6일

## 7. 구성원 역할 분담

조원	역할 분담
신다윗	ECG 데이터 분석 및 전처리, 특징 파악, 이상치 및 노이즈 처리, <b>Attention Mechanism</b> 적용한 <b>CNN</b> 모델 적용 및 고도화 시키기
안주현	딥러닝 <b>Base</b> 모델 설계 및 학습, 모델 성능 평가, <b>Vision Transformer</b> 적용한 <b>CNN</b> 모델의 고도화시키기
이재현	<b>Vision Transformer</b> 적용한 <b>CNN</b> 모델의 고도화시키기, 딥러닝 모델 경량화 및 모델 개선