



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

경량화 모델을 활용한 심방세동 분류 및
엣지 디바이스 적용

Classification of Atrial Fibrillation using light
weight model and application on its edge
device

울 산 대 학 교 대 학 원
의 과 학 과
박 이 슬

경량화 모델을 활용한 심방세동 분류 및
엣지 디바이스 적용

지도교수 주세경

이 논문을 공학석사 학위논문으로 제출함

2024년 8월

울산대학교 대학원
의과학과
박이슬

박이슬의 공학석사 학위논문을 인준함

심사위원 남기병 인

심사위원 주세경 인

심사위원 김준기 인

울산대학교 대학원

2024년 8월



목 차

국문요약	II
표 목차	IV
그림 목차	IV
식 목차	V
서론	1
1. 연구 배경	1
2. 자동 심방세동 분류	6
3. 인공지능과 딥러닝	8
4. 경량 딥러닝 모델과 온디바이스 AI	12
본론	15
1. 연구 방법	15
2. 연구 평가	27
3. 연구 결과	29
토의	32
한계 및 고찰	33
결론	34
참고문헌	35
영문요약 (Abstract)	38

국문 요약

딥러닝 모델은 성능을 높이기 위해서 모델의 파라미터와 레이어수를 늘리고 네트워크의 복잡성을 증가시키는 방식으로 발전했다. 또한 발전된 딥러닝 모델을 의료 인공지능 분야에서 활용한 사례가 점진적으로 증가했다. 이제는 딥러닝 모델을 컴퓨터나 클라우드에서만 작동시키는 게 아니라 온디바이스 형태로 하드웨어에서 구동시키는 시대가 도래하였다. 기업들에서도 온디바이스 기기 적용하기 위한 움직임이 증가하는 추세이다.

온디바이스는 기기 자체에서 모델을 실행할 수 있으므로, 빠른 응답 속도와 저전력, 저비용, 개인정보 보호에 쉽다. 또한 클라우드나 서버에서 모델을 작동시켰을 때 가지는 처리속도나 보안성과 같은 단점들을 해결해준다. 그러나 온디바이스 인공지능은 하드웨어의 제약으로 인해 전력 사용량, RAM 사용, 저장공간, 계산력 등을 고려하여 설계해야 한다. 모델을 경량화 할 때 더 적은 처리 시간으로 더 많은 처리 요청 수가 가능하다.

심장 질환은 우리나라에서 3대 사망 질환 중 하나이다. 이 중 심방세동은 심방의 수축이 소실되어 불규칙하게 수축하는 상태로 빈번하게 발생하는 부정맥의 한 종류이다. 본 연구에서는 심방세동 환자의 홀터 심전도 데이터를 이용하여 학습과 평가를 진행하고자 한다.

Physionet 에서 제공하는 오픈 데이터 형식의 MIT-BIH Atrial Fibrillation Dataset을 이용한다. Interpatient와 Inpatient 두 가지의 패러다임을 진행하기 위해서 학습 데이터 세트를 두 그룹으로 생성한다. 그리고서 각 데이터 세트에서 RR 간격을 이용하여, 1개의 비트(beat)로 segmentation을 진행한다. 비트 단위로 나눈 데이터를 이용하여 심방세동과 심방세동이 아닌 것으로 분류하는 이중분류 학습을 진행한다. 학습은 정답이 있는 데이터를 학습시켜 결과에 대한 예측을 만들어내는 방법인 지도학습 방식으로 진행한다. 잔차 연결을 이용한 Resnet 기반의 총 3가지 모델을 이용해 지도학습 방법으로 학습을 진행하였다. 학습한 모델은 하드웨어 장치인 젯슨 나노 오리노에 이식하여 결과를 평가한다. 하이퍼파라미터는 학습율은 0.0001, 배치 사이즈는 128, 손실함수로는 binary_cross_entropy를, 마지막 밀집층의 활성화 함수는 softmax를 사용하였다.

연구 결과는 분류 모델의 대표적인 평가 지표인 Precision (정밀도), Recall (재현율), F1 score (F1 점수), Accuracy (정확도), Area Under Curve (곡선 아래의 면적)를 이용하여 평가했다. 학습과 평가에 같은 환자의 레코드가 섞인 Inpatient 데이터 세트의 결과가 겹치는 환자의 데이터가 없는 Outpatient 데이터 세트를 이용하여 학습한 결과보다 월등히 좋은 값을 기록했다. Inpatient에서 제일 좋은 결과를 보인 성능은 모델에서 F1 점수가 0.9763이었다. 반면에 Outpatient에서 제일 좋은 결과를 보인 모델은 F1 점수가 0.8589이었다. 이를 통해서 심방세동 분류 모델이 온디바이스 기기의 실행을 실질적으로 확인 할 수 있었다.

본 연구가 가지는 한계점은 한 종류의 데이터셋으로만 학습했다는 점과 경량화 모델을 한 종류의 하드웨어에서만 평가했다는 점과 경량화 방식 중 하나만 선택한 점이다. 이러한 한계점에도 불구하고 장시간 심전도 기록에서 심방세동을 찾는 과정에서 소요되는 시간과 비용을 감소시킬 수 있고, 경량화 모델을 이식한 온디바이스 기기를 이용해 장소와 환경에 구애받지 않고, 어디서든 모델을 실행할 수 있다는 측면에서 편리성과 실용성이 갖춰졌으므로 임상적으로도 의의가 있는 것으로 판단된다.

표 목차

- 표1. 최근 5년('18' ~ 22년) 연령별 심장 질환 환자 수 (환자비율)
- 표2. MIT- BIH Atrial Fibrillation Dataset
- 표3. 학습에 쓰인 모델
- 표4. 젓슨 오린 나노 하드웨어의 스펙
- 표5. 3가지 모델을 통해 평가한 정밀도, 재현율, F1 - 점수

그림 목차

- 그림1. 정상인과 심방세동 환자의 심장의 리듬 변화와 심전도 비교
- 그림2. 심방세동의 임상발현과 관련 결과
- 그림3. 심전도
- 그림4. 인공지능과 머신러닝과 딥러닝
- 그림5. 퍼셉트론의 기본 단위 활성화함수
- 그림6. 활성화함수
- 그림7. 알고리즘 종류에 따른 손실함수
- 그림8. 연도에 따른 LLM 모델의 발현 및 파라미터 개수 비교
- 그림9. 오른쪽 위부터 라즈베리파이, 아두이노, 젓슨 나노 및 젓슨 오리노
- 그림10. 온디바이스 인공지능과 클라우드 서버의 차이점
- 그림. 주요 기업들의 온디바이스 인공지능 준비 현황 및 사용한 LLM 모델
- 그림12. QRS 복합체가 감지된 심전도, Record 04015
- 그림13. QRS 복합체가 감지된 심전도, Record 04076
- 그림14. QRS 복합체가 감지된 심전도
- 그림15. 전체적인 연구 흐름
- 그림16. Raw ECG와 필터링된 ECG
- 그림17. RR간격이 표시되어있는 심전도
- 그림18. 데이터 분배 비율
- 그림19. skip connection
- 그림20. 15개 레이어 모델의 구조
- 그림21. 21개 레이어 모델의 구조
- 그림22. 25개 레이어 모델의 구조
- 그림23. 젓슨 오린 나노 하드웨어 장치

그림24. confusion matarix (혼동행렬)

그림25. ROC curve

그림26. Interpatients 데이터 세트로 학습한 모델별 결과.

그림27. Inpatients 데이터 세트로 학습한 모델별 결과.

식 목차

식 1 평가지표 계산 공식

서론

1. 연구배경

심장 질환은 매년 전 세계적으로 약 1,790만 명의 목숨을 앗아가는 주요 사망원인이다. [1] 대한민국 통계청에서 제공한 2022년 사망원인 통계에서 3대 사망원인 중 심장 질환이 포함되었고, 인구 10만 명당 65.8명을 기록했다. 대한민국 연령별 심장 질환 환자 수는 전 연령대에서 지속적으로 증가 중이다. [2] 표1에서 2018년부터 2022년까지의 심장 질환 환자 수의 소계를 확인했을 때, 매년 꾸준히 상승하였고 연평균 증감률은 평균 4.8%이다. 또한 10~19세 그룹과 20~29세 그룹의 연평균 증감률이 약 9%로 젊은 층에서 증가하는 양상이다.

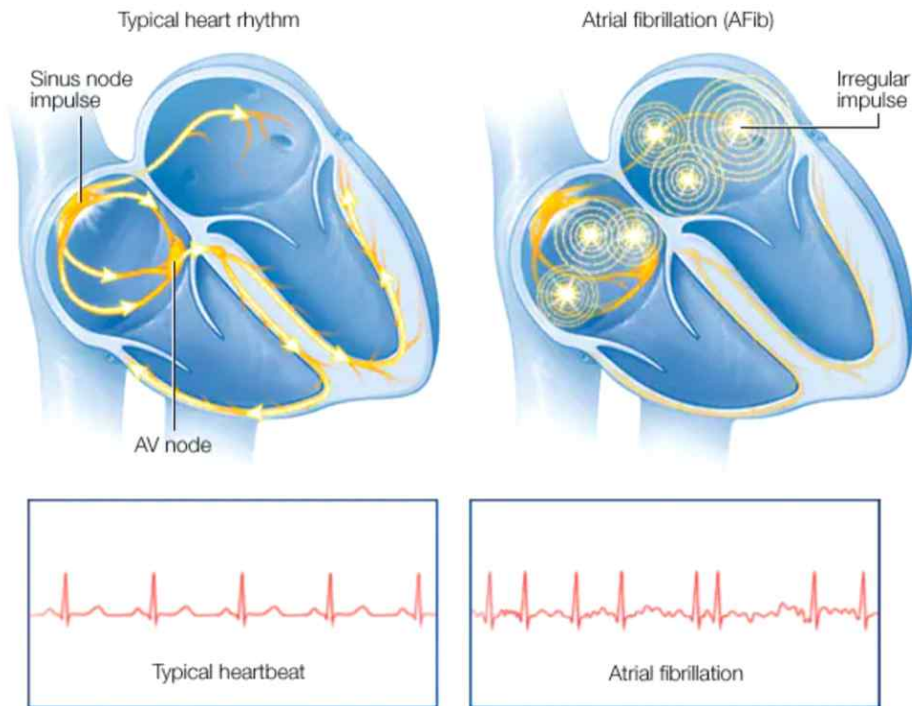
(단위: 명, %)

구분	2018년	2019년	2020년	2021년	2022년	환자비율 증감률	
	환자 수 (환자비율)	환자 수 (환자비율)	환자 수 (환자비율)	환자 수 (환자비율)	환자 수 (환자비율)	연평균	'18년 대비 '22년
소 계	1,529,537 (2.95)	1,610,231 (3.11)	1,625,042 (3.14)	1,785,612 (3.46)	1,833,320 (3.56)	▲ 4.8	▲ 20.8
10대 미만	6,610 (0.15)	6,740 (0.16)	5,167 (0.13)	5,619 (0.15)	6,024 (0.17)	▲ 2.6	▲ 11.0
10~19세	10,210 (0.20)	10,179 (0.21)	8,766 (0.18)	13,123 (0.28)	13,153 (0.28)	▲ 8.9	▲ 40.5
20~29세	22,802 (0.33)	23,913 (0.35)	22,627 (0.33)	37,460 (0.56)	30,215 (0.47)	▲ 9.0	▲ 40.9
30~39세	41,823 (0.58)	42,453 (0.60)	39,846 (0.58)	54,433 (0.81)	48,458 (0.73)	▲ 6.2	▲ 27.3
40~49세	111,926 (1.32)	111,909 (1.33)	106,762 (1.29)	122,307 (1.50)	117,275 (1.45)	▲ 2.5	▲ 10.2
50~59세	288,835 (3.35)	293,879 (3.39)	280,838 (3.25)	292,252 (3.38)	289,633 (3.36)	▲ 0.1	▲ 0.3
60~69세	437,248 (7.35)	463,996 (7.35)	477,332 (7.08)	522,373 (7.29)	545,854 (7.37)	▲ 0.1	▲ 0.3
70~79세	440,337 (12.59)	464,299 (12.91)	472,084 (12.77)	497,816 (13.40)	517,731 (13.53)	▲ 1.8	▲ 7.5
80대 이상	235,675 (13.50)	263,506 (13.98)	281,281 (14.04)	313,053 (14.85)	349,035 (15.47)	▲ 3.5	▲ 14.6

* 환자 수는 동일인의 중복을 제거한 값으로, 단순 합산할 경우 중복이 발생할 수 있음
 * 연령은 명세서에 기재된 생년월일(만(滿) 나이)을 계산한 연령으로, 진료시점에 따라 중복집계가 될 수 있음
 * 환자비율은 연령별인구 대비 환자 수 비율

표1. 최근 5년('18 ~ 22년) 연령별 심장 질환 환자 수 (환자비율)

(출처: 의협신문, 2023)



© MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH. ALL RIGHTS RESERVED.

그림1. 정상인과 심방세동 환자의 심장의 리듬 변화와 심전도 비교

(출처: <https://www.mayoclinic.org/diseases-conditions/atrial-fibrillation/symptoms-causes/syc-20350624>)

심장질환 중 하나인 심방세동을 가지는 환자와 정상인의 심장은 구별된다. 그림1에서 심장은 총 4개의 영역으로 나뉜다. 심장의 윗부분은 우심방과 좌심방이고, 아랫부분은 우심실과 좌심실이다. 정상인의 심장은 우심방 위쪽의 동방결절(Sinus node)에 있는 세포들이 전기 신호를 생성하여 보낸다. 전기 신호는 심방에서 방실결절(Atrioventricular node)로 이동한다. 이후 심실로 이동하여, 심실이 수축했다가 뿜어내며 혈류를 흘려 보내는 방식이다. 심방세동 환자의 심장에서는 심방의 전기신호에 이상이 생겨 원래대로 수축하지 못하고, 불규칙하게 진동하는 상태가 된다. 심실로 이상한 전기신호들이 가는 것을 방실결절이 막지 않으므로 심장 박동은 더 빠르고 불규칙적으로 뛰게 된다. 이는 부정맥이라고도 한다. 휴식 상태에서 정상인의 심장은 분당 60번에서 100번까지 뛸 수 있지만, 심방세동의 경우 분당 100번 이상 뛰고, 각 박동이 불규칙하다.[3]

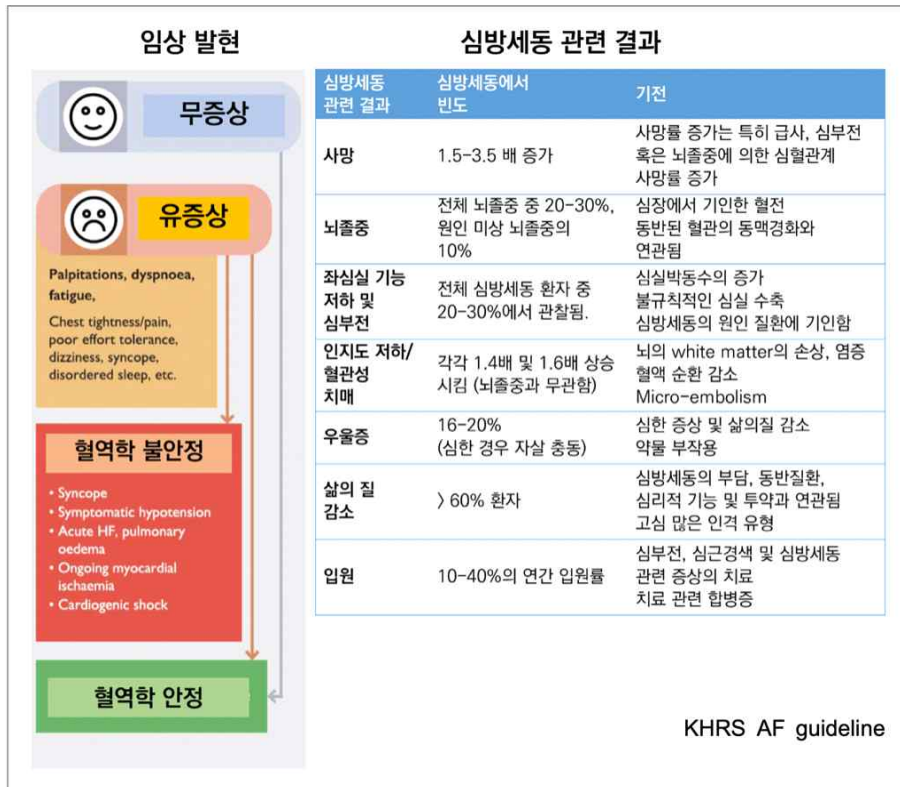


그림2. 심방세동의 임상발현과 관련 결과
(출처: 2021 대한 부정맥 학회, 심방세동 진료지침, p18)

심방 세동은 일시적인 것과 지속적인 것으로 나뉜다. 대부분 심방세동의 증상은 발현되지 않는다. 발현되는 증상에는 빠르게 두근되는 심장 박동과 호흡곤란, 어지러움이 있다. 이는 심장이 수축할 때 뿜어내는 혈액의 양인 심박출량이 감소하기 때문이다. 정상 심장에서 심박출량에 심방의 수축이 20% ~ 30%를 차지하지만, 심방 세동의 경우 심방의 수축이 사라져 심박출량이 감소한다. 또한 맥박수가 과도하게 빨라 심장에 혈액을 채울 시간이 부족하므로 심박출량이 감소한다. 또한 심방세동으로 인해 심장 내 혈전이 발생할 경우 뇌졸중으로 이어질 수 있다. 심방 세동 자체가 생명에 위협이 되는 질환은 아니지만, 뇌졸중이나 심부전등의 다른 질환들을 야기시킨다. 그림2에서 심방세동 환자가 느끼는 다양한 증상들로 인해 삶의 질이 저하되고, 합병증으로 인해 사망률이 증가한다. 그러므로 조기에 진단하여 적절한 치료를 시행하며 관리해야한다.

심방세동을 진단하기 위해서는 기본적으로 심전도 검사를 시행한다. 심전도(ECG : Electrocardiogram)는 피부에 전극을 붙이는 비침습적인 방식

으로 심장에서 발생하는 전기 신호를 측정하여 주기적인 펄스 패턴인 심장에 대한 정보를 획득한다. 심전도 검사는 검사 장비에 따라서 표준 12 리드 심전도와 운동 부하 심전도와 활동 중 심전도(Holter monitoring)가 있다.

심전도는 약 10초 동안 전기 신호를 기록하기에 자주 발생하지 않는 부정맥은 검사 도중 나타나지 않으면 찾을 수 없으므로, 심방세동의 경우 홀터 검사를 통해 24시간에서 한 달여 동안 심전도를 기록한다. 검사를 통해 획득한 심전도 신호의 파형은 P파, QRS 복합체, T파로 구성되고, 이를 심장박동의 한 주기인 비트(beat)라고 부른다. (그림 3) 한 비트에서 QRS 복합체는 최대 진폭을 가진다. 그리고 P파는 심방의 탈분극을 나타내고, QRS 복합체는 심실의 탈분극을, T파는 심실의 재분극을 나타낸다. 심전도를 통해서 심장의 크기와 위치, 심장의 손상 여부 및 손상 정도, 심장을 조절하는 장치나 약물에 대한 반응을 알 수 있어 임상적으로 유용하다.

[4]

심전도 신호에서 P파의 부재와 불규칙한 심실 속도는 심방세동과 연관되어 있을 수 있다. 심실의 속도는 비트의 QRS 복합체와 다음 QRS 복합체 사이의 거리인 RR 간격의 비율을 통해 알 수 있다. 그림1에서 정상인의 심전도는 RR 간격이 일정하지만, 심방세동 환자의 심전도는 RR 간격이 길거나 짧은 양상을 보이며 불규칙하다.

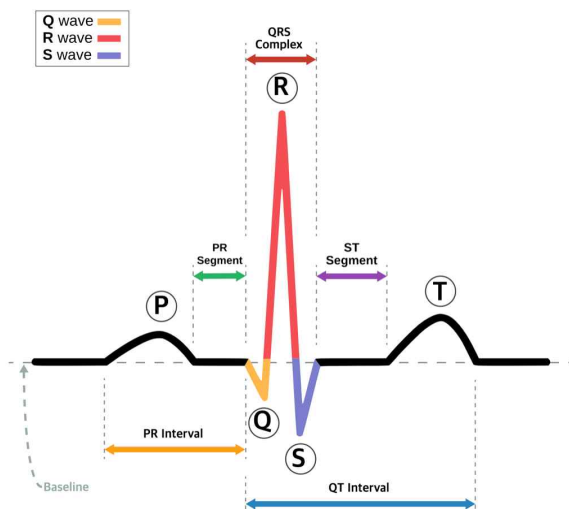


그림 3. 심전도

(출처: https://en.wikipedia.org/wiki/Electrocardiography#/media/File:Sinus_RhythmLabels.svg)

임상의가 심전도 검사의 결과를 분석하고 관찰하는 기존의 방식은 주로 개인의 임상 경험이나 분석을 위한 기존 배경지식에 크게 의존한다. 이는 임상의들에게 번거롭고 의료 자원을 많이 필요로 하여 부담을 줄 수 있다. 또한, 진단 결과가 심장내과 의사의 주관성에 크게 영향을 받기도 한다. 머신 러닝은 이러한 문제를 해결해 줄 수 있다. 자동 분류 알고리즘을 개발하면 의사의 작업량을 줄일 수 있다. 긴 시간 동안 기록된 심전도 데이터를 진단하는 데 걸리는 시간을 단축하여 진단의 객관성과 일관성을 높일 수 있다. 많은 연구에서 심전도 신호의 이상을 감지하기 위해 다양한 특징 추출 및 분류 기술을 사용하는 시스템들을 제시하였다.

2. 자동 심방 세동 분류

자동 심방세동 분류는 최근 몇 년 동안 큰 발전을 이루었다. 심전도 기반의 오픈 데이터셋을 이용하여 다양한 모델들이 출시되었다. 초기에는 심전도 신호에서 주요한 특징을 추출하고, 이를 기반으로 머신 러닝 모델을 훈련 시키는 방식이었다. 서포트 벡터 머신, 선형 회귀, 머신 러닝 기법, 선형 분류, K-NN(K-Nearest neighbor), XGBoost등의 방법이 널리 사용되었고, 이후에는 딥러닝 기반의 방법들이 주목을 받게 되었다.

딥러닝 방식은 합성곱 신경망(Convolutional Neural Network, CNN)과 순환신경망(Recurrent Neural Network, RNN)과, 심층 신경망(Deep Neural Network, DNN), 장단기메모리(Long Short Term Memory, LSTM)등의 모델이 있다. CNN은 자동으로 중요한 특징을 추출하고 많은 양의 데이터가 있을 경우 더 잘 일반화 하고, RNN은 시계열 데이터를 효과적으로 처리한다. DNN은 여러 층을 통해 복잡한 패턴을 학습하고, 비선형 관계를 모델링하는데 뛰어나다. LSTM은 장기 의존성을 학습할 수 있어서 긴 시간동안의 시계열 데이터를 효과적으로 처리한다.

저자의 연구[5]에서는 1차원의 심전도 신호 데이터를 CNN과 RNN 기반의 LSTM을 사용하여 부정맥을 판별하였다. 공간 특성 매핑은 합성신경망의 층에서 추출되고, 이러한 매핑의 시간에 따른 변화는 이어지는 LSTM 모델의 층에서 수집된다. 심장의 이벤트가 발생하거나 수면 단계를 분류하기 위한 알고리즘에서는 2차원 합성 신경망 모델을 사용하는 것이 효과적이다. 이미지 형태의 2차원 데이터를 활용할 때 CNN은 여러 크기의 필터를 사용하며 공간적 특성을 분석하고 인식하는데 뛰어나다.

이러한 알고리즘은 데이터를 이용해 학습하고, 새로운 데이터를 추론하는 방식에 따라 지도 학습, 비지도 학습, 강화 학습으로 분류할 수 있다. 지도 학습은 학습 데이터와 라벨이 둘 다 제공되는 경우이고, 비지도 학습은 라벨이 제공되지 않는 경우이고, 강화학습은 주어진 환경에서 보상을 최대화하기 위해 행동하는 에이전트를 훈련하는 경우이다. 심방세동의 분류도 이 3가지의 학습 형태 중 하나를 선택하는 식으로 발전되었다.

저자의 연구[6]에서는 지도학습 방식으로 표준 12 리드 심전도 데이터를 CNN 모델로 학습하여 심방세동을 포함한 10개의 클래스로 분류하였다.

또한 저자의 연구[7]에서는 비지도 학습 방식으로 다중지점잔차망 (Multi-branch Residual Network, MRN) 모델과 적대적 생성 모델 (Generative Adversarial Model, GAN)을 학습하여 5개의 클래스로 분류하였다. 저자의 연구[8]에서는 강화학습 방식으로 CinC(PhysioNet/Computing in Cardiology Challenge)에서 제공하는 다양한 데이터셋을 분석하고 5개의 심전도 비트로 분류하였다.

이와 같이 많은 연구에서 심방 세동 분류를 진행하였지만, 여전히 개선할 부분이 존재한다. 첫째로, 대부분의 연구에서 inpatient 방식으로 수행되었다. 학습과 테스트용 데이터를 분배하는 방식을 통해 interpatient와 inpatient 패러다임으로 나뉜다. inpatient의 경우 동일한 환자의 데이터를 학습과 테스트에 사용하므로 환자의 개인적인 특성이나 변화를 더욱 정확히 반영할 수 있다. 따라서 환자 개개의 치료 예방 전략을 수립하는데 유용하다. 반면에 interpatient로 학습한 모델은 다양한 환자의 데이터를 사용하여 학습하므로 모델의 일반화 되는 능력이 향상된다. 이는 모델이 처음 보는 환자의 데이터에 대해서도 상대적으로 안정적인 예측을 할 수 있게 한다.

3. 인공지능과 딥러닝 (Artificial Intelligence and Deep Learning)

인공지능은 딥러닝보다 상위 개념이고, 인간의 학습, 추론, 지각 능력을 모방하기 위해 컴퓨터 프로그램을 이용하여 실현한 기술을 의미한다. 머신러닝은 사람이 훈련 데이터 세트에서 특정 패턴을 추출할 수 있도록 방법을 알려주고, 그 이후에 컴퓨터가 스스로 데이터의 특징을 분석하여 출력 값을 도출하는 방식으로 문제를 해결한다. 그러나 딥러닝은 인간이 개입해야 했던 패턴 추출 작업이 생략된다. 이는 훈련 데이터세트를 신경망에 입력하였을 때, 컴퓨터가 스스로 데이터를 이용해서 특징을 추출하고 패턴을 분석하는 학습이 가능해진 것이다.

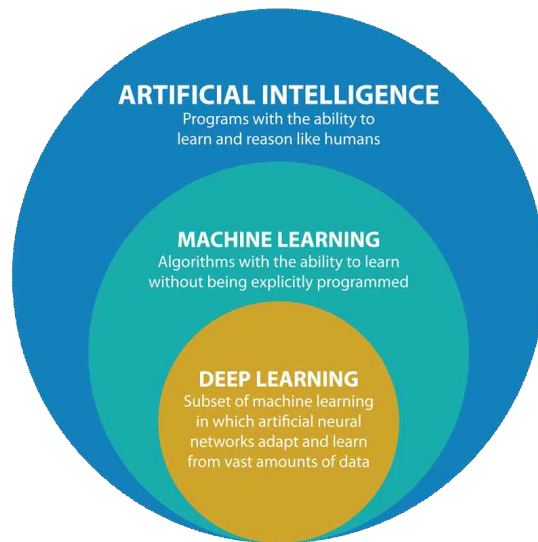


그림4. 인공지능과 머신러닝과 딥러닝

(출처: <https://www.thetechedvocate.org/the-two-main-barriers-against-deep-learning/>)

딥러닝에서 모델이 데이터를 처리하고 학습하는 방식은 신경망의 기본 단위인 퍼셉트론으로 설명된다. 다수의 신호를 입력받아 하나의 신호로 출력한다. 다중 퍼셉트론은 입력층과 출력층 사이에 은닉층을 추가하여 모델의 성능을 증가시킨다. 은닉층은 하나 이상 존재 할 수 있고, 층이 증가할 수록 입력데이터의 복잡한 특징을 잘 학습할 수 있다. 은닉층에서는 입력받은 신호들을 비선형 함수인 활성화 함수를 통해서 처리한다. 비선형 함수를 이용하는 까닭은 신경망의 표현성을 높이고, 선형 함수가 들어간 층을 여러 개 쌓았을 때 단 하나의 층으로 이를 대체할 수 있기 때문이다. 퍼셉

트론에서 입력 데이터를 넣으면 입력값에 가중치를 곱하고 바이어스를 더한다. 이 값을 활성화 함수에 넣어 출력값을 얻는다. 이렇게 얻은 출력값이 실제 정답과 얼마나 차이가 있는지 계산한다.

활성화 함수는 선형적인 값을 비선형적으로 바꿔주고, 예시로 sigmoid 함수, tanh 함수, ReLU 함수 등이 있다. 또한 모델의 마지막 밀집 계층에서 손실함수를 이용하여 출력값과 정답 값 사이의 차이를 평가한다. 손실함수는 분류하려는 문제의 종류에 따라 달라질 수 있다. (그림 6) 두 개의 클래스를 분류하는 이진 분류의 경우 이진 크로스 엔트로피 로스 방법이나 MSE(Mean squared error), RMSE(Root mean squared error), Binary cross entropy 등을 쓸 수 있다. 2개 이상의 클래스를 분류하는 다중 분류 문제에서는 categorical cross entropy 함수를 사용할 수 있다.

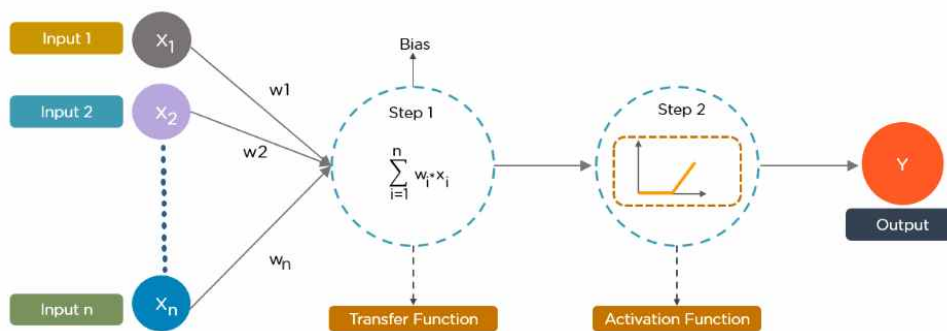
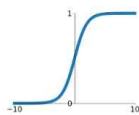


그림5. 퍼셉트론의 기본 단위 활성화함수

Activation Functions

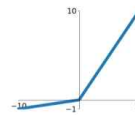
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



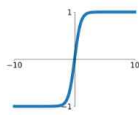
Leaky ReLU

$$\max(0.1x, x)$$



tanh

$$\tanh(x)$$

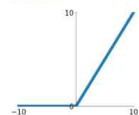


Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

ReLU

$$\max(0, x)$$



ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$

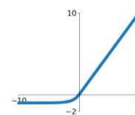


그림6. 활성화함수

(출처: <https://medium.com/@shrutijadon/survey-on-activation-functions-for-deep-learning-9689331ba092>)

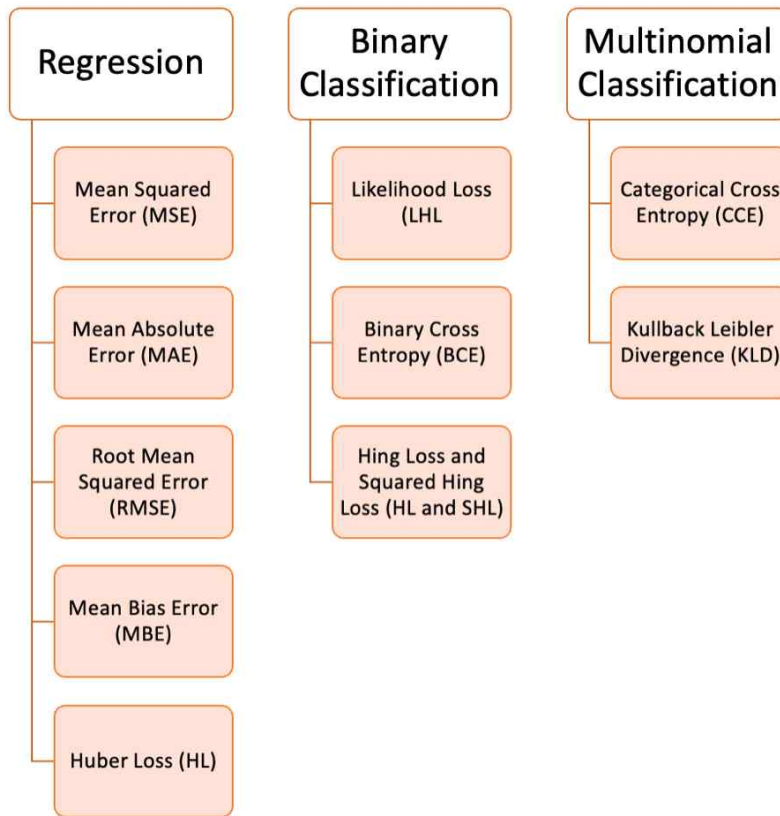


그림7. 알고리즘 종류에 따른 손실함수

(출처:<https://pub.towardsai.net/10-commonly-used-loss-functions-explained-with-python-code-59967e1f3c8d>)

현재 합성 신경망 기반의 간단한 딥러닝 모델부터 모델의 층이 여러 개 존재하고 대용량의 데이터셋을 한 번에 처리하는 거대한 딥러닝 모델까지 개발되었다. 거대한 개수의 파라미터와 다량의 계층을 가지는 모델에는 대형 언어 모델(Large Language Model, LLM)과 대형 멀티 모델(Large Multimodal Model, LMM)이 있다. 2024년 5월에는 GPT-4o가 출시되었고, 이 모델의 파라미터의 개수는 기존의 GPT-3가 가지는 1,750억 개보다는 많을 것으로 예상된다. (그림 8) 모델의 성능은 학습에 사용되는 데이터에 영향을 받으므로 많은 양의 데이터가 요구된다. [9] 데이터가 많을수록 모델이 학습할 수 있는 예시가 증가하여 과적합 가능성이 줄어들 수 있다. 큰 데이터셋을 이용한 대형 모델들은 학습시키는데 소요 시간이 오래 걸리고, 학습에 많은 양의 그래픽 처리 장치 (Graphic Processing Unit, GPU) 가 필요하다. [10]

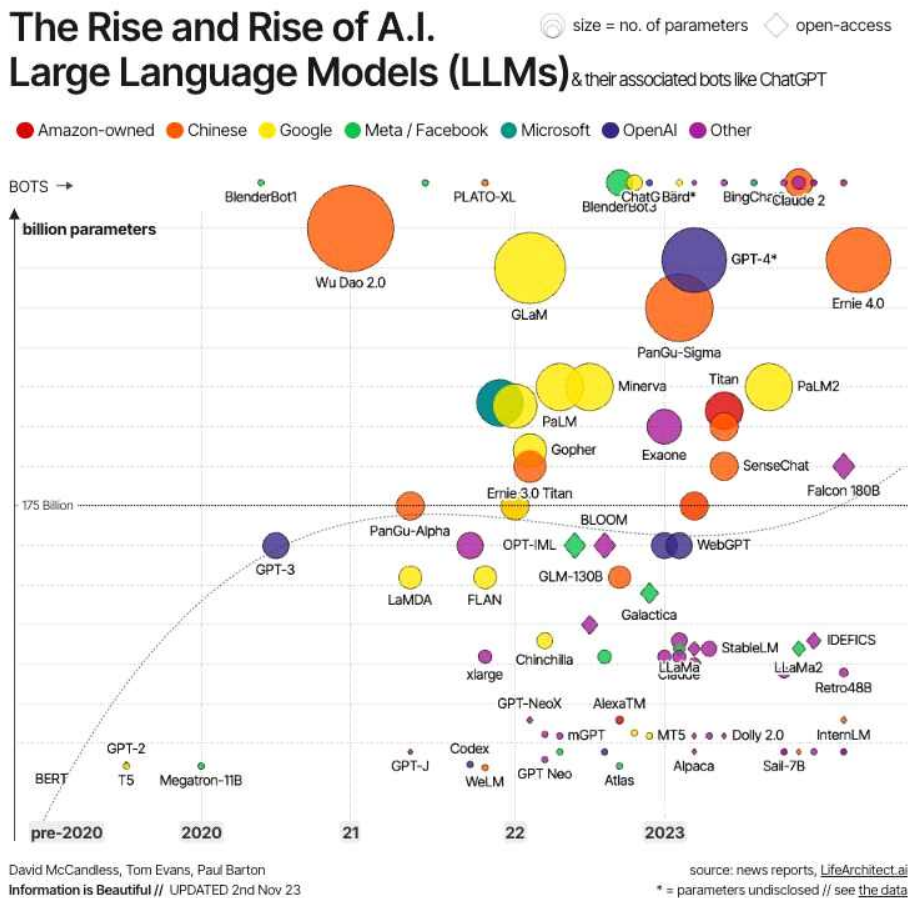


그림8. 연도에 따른 LLM 모델의 발현 및 파라미터 개수 비교

(출처: https://www.linkedin.com/posts/a-banks_the-development-of-ai-large-language-models-activity-7089944984501506048-zXhP/)

4. 경량 딥러닝 모델과 온디바이스 AI

대용량 모델을 상용화하여 제한된 컴퓨팅 자원에서 실행하는 것은 비실용적이다. 그러나 딥러닝 모델을 경량화하여 개발하면 효율성과 소형화를 동시에 달성할 수 있다. 경량화된 모델은 속도와 지연시간이 짧으므로 저전력 휴대폰, 임베디드 시스템 및 사물 인터넷 장치에 이상적이다.[11]

이러한 경량화 방법에는 양자화, 가지치기, 지식 증류, 모델 압축, 모델 구조 최적화, 신경망 구조 탐색(Neural Architecture Search, NAS), 파라미터 공유, 전이 학습 등이 사용된다. 경량 딥러닝 모델로 SqueezeNet [12], MobileNet, ShuffleNet, GhostNet [13]등이 개발되었다. 또한 ResNet과 같은 높은 성능을 내는 모델의 구조를 변경하여 경량화시키기도 한다.

경량화 모델을 하드웨어에 이식하면 온디바이스 AI를 구현할 수 있는데, 온디바이스 AI는 엣지 디바이스로도 불린다. 기존의 클라우드 방식은 데이터 수집과 전송에 많은 시간이 소요되고, 인터넷이 존재하는 환경에서만 작동하였다. 그러나 온디바이스 또는 엣지 디바이스는 클라우드 서버를 거치지 않고, 통신 연결 없이 자원이 제한된 하드웨어 안에서 자체적으로 정보를 처리한다. 이는 작업 속도를 향상하고, 대역폭과 저장공간의 문제도 해결해 준다. 또한 클라우드 기반 인공지능에서 거론되었던 보안 문제도 피할 수 있다.(그림 10) 따라서 보안이 전제되는 상황에서는 개인화된 인공지능으로 사용자의 개인 데이터 (나이, 활력징후, 심전도 등) 나 카메라, 자이로스코프 등 센서 데이터를 더욱 안전하게 다룰 수 있다. 이로써 미래 의학의 핵심 4P 중 개인 맞춤 의학도 달성할 수 있는 것이다.

최근 하드웨어 및 웨어러블 디바이스에 대한 수요가 증가하고 있다. (그림 11) 온디바이스 인공지능은 다양한 소형 하드웨어로 구현할 수 있다. (그림 9)온디바이스 인공지능은 개발 단계에서 하드웨어에 제한된 성능과 공간을 고려해야 한다. 매개 변수를 줄이면서도 성능의 저하를 최소화 하는 것이 온디바이스 인공지능 개발의 필수 요소이다.

온디바이스는 앞으로 활용 가능성이 무궁무진하기에 스마트폰, 웨어러블 기기, 가정용 스마트 기기 뿐만 아니라 더욱 다양한 분야에서 우리 생활과 밀접하게 나타날 것이다. 웨어러블 디바이스의 온디바이스가 의료 분야에

적용될 때 의사가 신속하고도 정확하게 진단을 내리는 데 도움을 줄 수 있다. 실시간 모니터링, 빠른 진단, 개인화된 치료, 조기 감지 등의 효과를 기대할 수 있다.

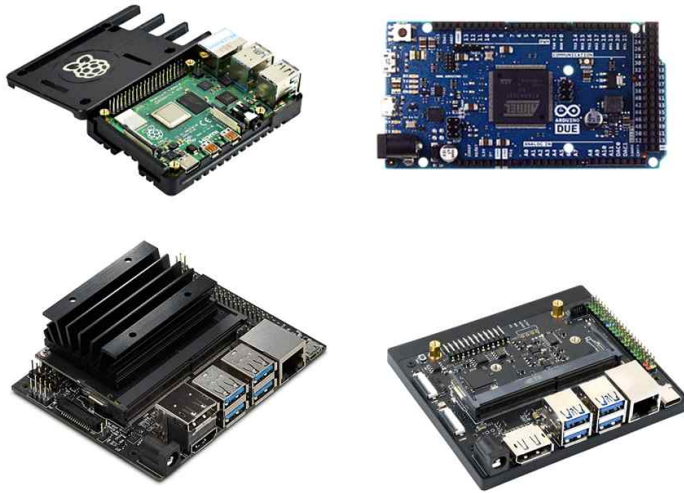


그림9. 오른쪽 위부터 라즈베리파이, 아두이노, 젯슨 나노 및 젯슨 오리노

온디바이스시와 클라우드서버시 장단점 비교

구분	클라우드서버시	온디바이스시
장점	대용량, 거대 규모 모델 운용을 통한 시 학습 및 추론 성능 극대화	빠른 응답속도, 저전력, 저비용, 개인정보 보호
단점	추론 정확도 중요(저전력화 는 상대적으로 후순위)	실시간 추론&저전력 중요

그림10. 온디바이스 인공지능과 클라우드 서버의 차이점

(출처: <https://www.mk.co.kr/news/culture/10911744>)

주요 테크 업체들의 LLM 및 온디바이스시 준비 현황

OEM	LLM	LLM 및 On-Device AI 준비 현황
 삼성전자	Gauss	자체 생성형AI인 가우스는 언어모델, 코드모델, 이미지모델 등 3가지 모델로 구성 이를 사내에서 활용 & 보완한 후 모바일에 탑재할 계획. 아직 자세한 사양이 공개되진 않음
 애플	Ajax	공식 LLM 서비스 발표 없음. 다만 블룸버그는 애플이 LLM R&D에 10억달러 투자를 보드 2024년 발표될 iOS 18에 Ajax를 지원 → 챗봇 서비스인 AppleGPT 출시할 것이라 전망
 화웨이	Pangu	자체 클라우드 LLM인 Pangu와 디바이스 내 AI버서인 Celia가 연결되어 있음 나이가, 8월 개발자 회의에서 자체 OS인 HarmonyOS 4에 LLM을 탑재할 것임을 발표
 샤오미	MiLM	2023년 4월 AI 모델 경량화를 위한 전담 LLM 연구팀을 설립. 2023년 연간 200억위안을 투자 모바일 비서 Xiao AI에 자체 LLM인 MiLM을 탑재, 초기 테스트 중(월간 1, 1억명 사용자)
 오포	AndesGPT	8월 초, 오포는 자체 LLM인 AndesGPT를 기반으로 AI버서인 Xiaobu를 외부 테스트 AndesGPT의 모델은 매개변수 1억, 3억, 10억개로 구분되며, 향후 스마트폰 등에 탑재
 아너	외부협력	Shanghai MWC 당시 LLM 대응을 위해 인터넷 서비스 업체와 협업하고 있다고 언급 바이두의 어니봇, 알리바바의 Tongyi Qianwen 등이 후보군이 될 수 있음

*자료: Digitimes, Bloomberg, 언론 정리, 하이투자증권 리서치본부

그림11. 주요 기업들의 온디바이스 인공지능 준비 현황 및 사용한 LLM 모델

(출처: <https://www.mk.co.kr/news/culture/10911744>)

본론

연구 방법

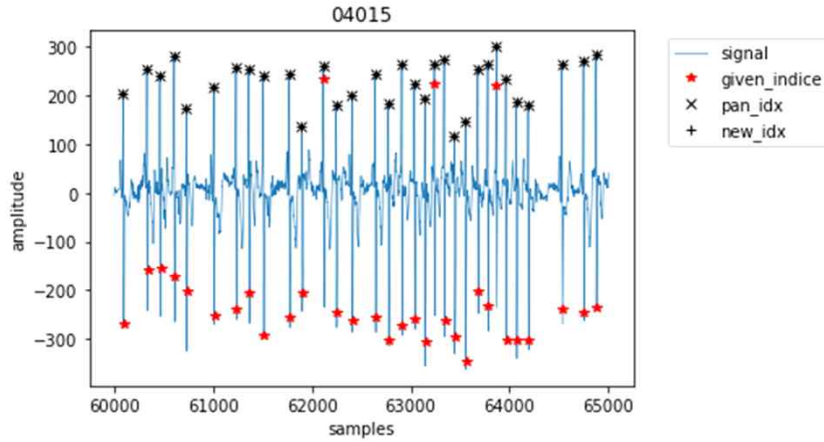


그림12. QRS 복합체가 감지된 심전도, Record 04015

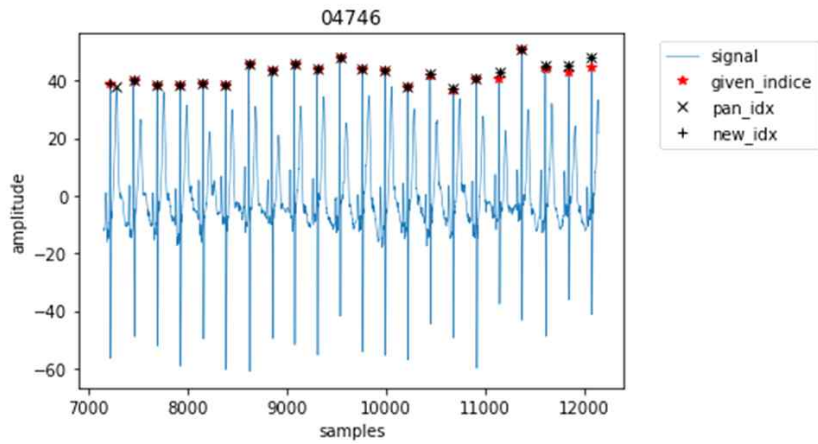


그림13. QRS 복합체가 감지된 심전도, Record 04076

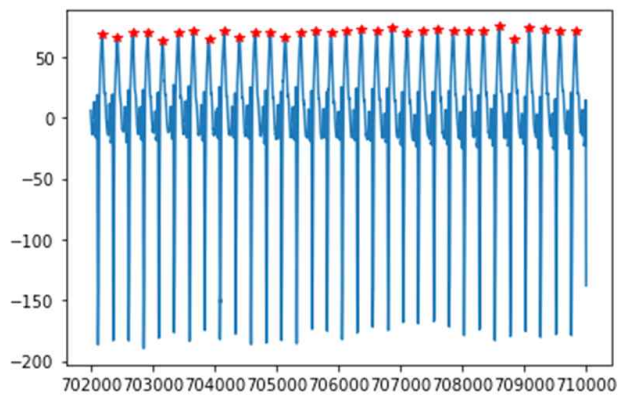


그림14. QRS 복합체가 감지된 심전도

(1) 데이터 세트

데이터세트는 Physionet 사이트에서 제공해 주는 MIT-BIH Atrial Fibrillation Database를 이용하였다.[14] Database는 보스턴 베스 이스라엘 병원(Boston's Beth Israel Hospital)에서 심방세동(대부분 발작성)이 있는 피험자들의 장기 심전도(Long term ECG) 데이터를 획득했다. 이 데이터 세트는 23개의 레코드로, 각 레코드는 10시간짜리인 두 리드의 심전도 신호를 포함하며, WFDB 형식으로 제공된다. 이 레코드는 12 bit resolution으로 1초당 250개의 심전도 신호 표본이 획득되었다. 주석은 AF(Atrial Fibrillation), AFL(Atrial Flutter), J(AV Junctional rhythm), N(Normal)으로 구성되어 있다. 각 주석은 누락되는 비트가 하나도 없이 라벨링 되도록 자동 탐지기를 이용하였다. 그런데 이 라벨은 각 피크와 가까운 위치의 피크를 잡는 형식으로 탐지 되어있기에, 더욱 정확한 피크의 위치 정보를 얻도록 판 톰킨스 알고리즘을 이용하여 얻은 라벨과 비교하여 사용하였다. (그림 12, 그림 13)

Database	Publicly Available	Release Year	No. Channels	Sampling Rate (Hz)	Duration	Subjects	Collection Place	No. of Records	Demographic Information
MIT-BIH Atrial Fibrillation Database (AFDB)	Yes	2000	2	250	10h	25	USA	23	Subjects are suffering from atrial fibrillation

표2 . MIT- BIH Atrial Fibrillation Dataset

판 톰킨스 알고리즘은 심전도에서 QRS 복합체를 감지하기 위해 쓰이는 방식이다. 판 톰킨스 알고리즘은 일련의 필터를 적용하여 급격한 심장 탈분극의 주파수 내용을 강조하고 배경 잡음을 제거한다. 그런 다음 신호를 제공하여 QRS contribute를 증폭하므로 QRS 복합체를 식별하는 것이 더 간단 해진다. 빠르고 효과적인 방법이라서 피크 감지 방식 중에서 널리 쓰이는 방식이다.

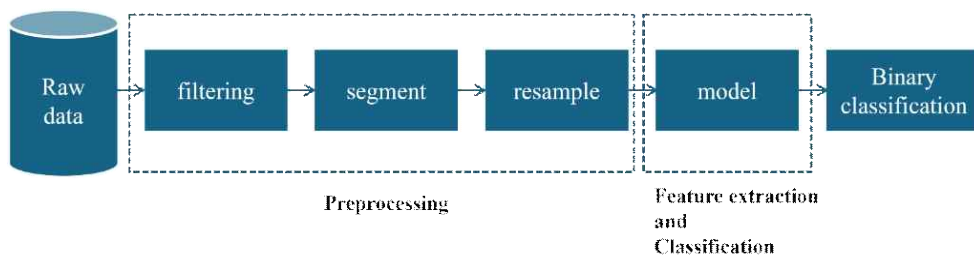
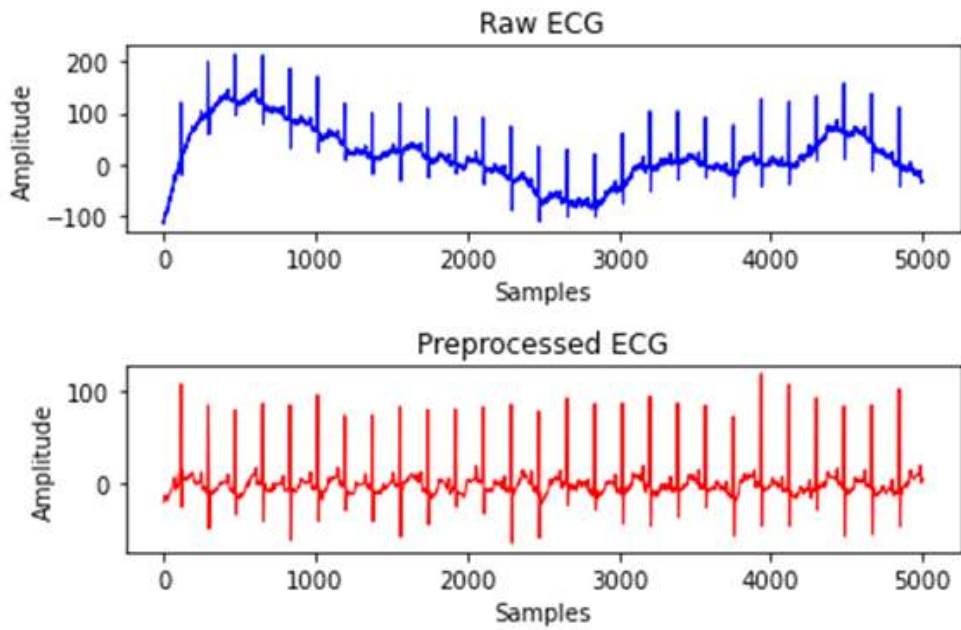
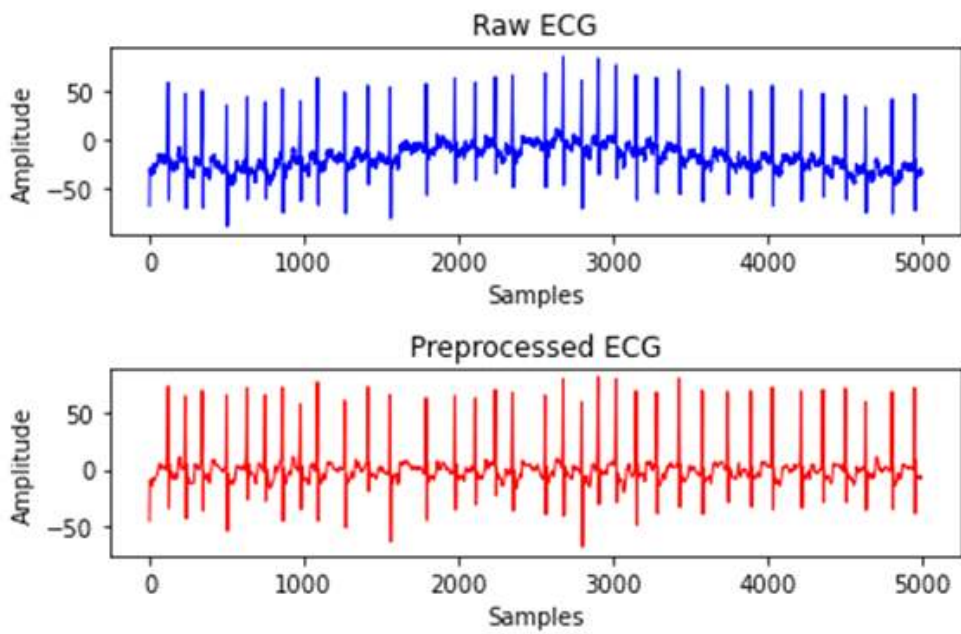


그림15. 전체적인 연구 흐름



(a)



(b)

그림16. Raw ECG와 필터링된 ECG

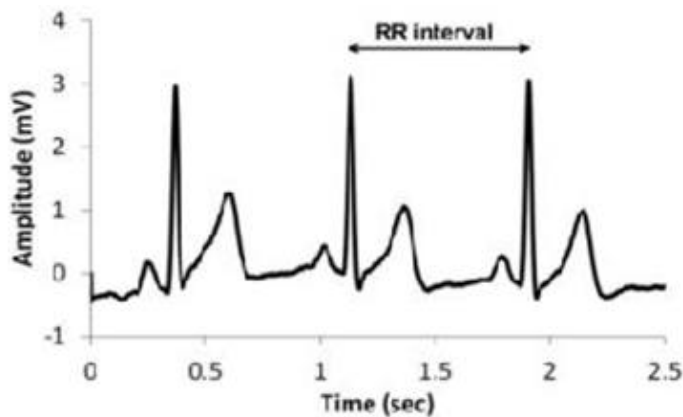


그림17. RR간격이 표시되어있는 심전도

두 개의 리드(lead) 중 'ECG 1'이라고 표시된 1번째 리드를 선택하였다. 데이터는 필터링을 거친 후 segmentation을 진행하였다. 전체 레코드는 비트 (beat) 단위로 학습할 수 있도록, segmentation을 진행하였다. 한 비트는 심장의 주기라고 할 수 있고, 심전도의 특징인 P파와, QRS 복합체와 T파가 모두 포함된 신호라고 할 수 있다. segmentation의 기준은 R peak와 R peak 사이의 간격인 RR interval(RR 간격)을 이용하였다.[15] 그 까닭은 심방세동의 경우 심박이 불규칙하므로, 동일한 길이로 segmentation을 진행할 때 비트의 모든 정보를 담지 못할 수 있기 때문이다. 이 연구에서는 (그림17)과 같이 연속된 3개의 피크를 이용하여, 첫 번째 피크와 두 번째 피크 사이의 거리(RR1)를 구하고, 두 번째 피크와 세 번째 피크 사이의 거리(RR2)를 구하였다. 그리고 첫 번째 RR 간격의 1/3에 해당하며 두 번째 피크와 가까운 쪽을 비트의 시작점으로 잡았다. 두 번째 RR 간격의 2/3 지점을 비트의 마지막 점으로 잡았다. 심방세동은 RR 간격이 일정하지 않기에 segmentation이 완료되었을 때 비트 간의 길이가 달랐으므로, 길이를 300개의 표본으로 보간법을 이용하여 oversampling을 해주었다. 모든 비트 (beat)가 보간이 된 R 피크의 위치가 동일한 지점에 있게 하도록, R 피크를 기점으로 두 부분으로 나누어 보간하였다. R 피크 앞부분은 100개의 표본으로, R 피크 뒷부분은 200개의 표본으로 늘린 후 합쳤다.

심전도는 전위 차가 아주 작은 신호이고, 작은 변동 사항에도 영향을 크게 받아 잡음이 많은 생체 신호이다. 기저선 변동, 전원 간섭, 전력선 간섭,

근잡음 등의 잡음으로 인해 신호가 오염될 수 있다. 기저선 변동의 경우 저주파수의 범위에 해당하고, 대체로 0.05Hz에서 0.5Hz 사이의 주파수를 가진다. 전원 간섭의 경우 고주파 잡음 범위에 해당하고, 대체로 50Hz에서 60Hz 사이에서 발생할 수 있다. 근 잡음은 주로 고주파 범위에 해당하고, 10Hz에서 100Hz 사이의 주파수를 가질 수 있다.

딥러닝 모델 학습의 효율성을 높이고, 딥러닝 모델의 연산 복잡성을 줄이기 위해 신호 처리 과정을 거쳐 심전도의 잡음을 제거했다. 신호 처리는 여러 개의 필터링을 이용하여 수행했다. 평균값 제거(remove the mean of signal), 대역통과 필터(bandpass filter), 이산 웨이블릿 변환(discrete wavelet transform)을 사용하였다. 평균값 제거를 통해서 데이터값들이 0에 수렴하도록 만들었다. 역전파 중에 계산된 그래디언트가 서로 다른 기능을 거쳐 일관되게 유지되어 학습 중에 속도가 빠르게 수렴될 수 있다. 이를 통해 심층 신경망 모델이 빠르고 안정적으로 훈련할 수 있다. 0.5Hz에서 50Hz 사이로 대역 통과 필터를 적용하였다. 대역통과 필터는 특정 범위 내의 주파수를 통과시키고, 그 범위 밖의 주파수를 막는 역할을 한다.

심전도는 비정상(non-stationary) 신호로 간주되는 생체신호로 잡음 제거 작업이 까다롭다. 이러한 비정상 신호 처리에 효과적인 방법은 웨이블릿 변환이다. ECG 신호는 주성분(고주파 요소)과 부성분(저주파 요소)으로 나뉜다. 분해할 때 Haar, Daubechies, Biorthogonal, Coiflets, Symlets, Morlet, Mexican Hat, Meyer 등과 같은 웨이블릿을 사용할 수 있다. 이 연구에서는 MIT-BIH arrhythmia 데이터세트에 웨이블릿을 적용한 저자의 논문 [16]처럼 주성분을 더욱 정확하게 제공해 주는 'Db4' 파형을 선택하였다. 이 파형은 QRS 복합체와 유사성을 가지고, 에너지 스펙트럼이 저주파에 집중되어 있다. 즉, 심전도의 특징을 추출하기 위해 'Db4'를 선택했다. 심전도 신호에 이산 웨이블릿 변환을 적용하여 신호를 각기 다른 주파수로 분해하고, 부성분에 대해 임계값을 가지고 soft thresholding을 진행하였다. 그리고 주성분과 임계값이 적용된 부성분과 같이 신호를 복원시켰다. 이 과정에서 또한 'db 4' (daubechieswavelet) 5차를 이용하였다. 모든 필터링 과정은 수집된 데이터를 가공하기 전과 후로 나타낸 것으로 볼 수 있다. (a)에서는 기저선 변동이 현저히 줄어든 것을 관찰할 수 있으며, (b)에서는 고주파 잡음들이 현저히 줄어든 것을 관찰할 수 있다.(그림 16)

(2) 데이터 그룹화

본 연구에서는 Interpatient와 Inpatient인 두 개의 패러다임으로 진행했다. MIT-BIH Atrial Fibrillation 데이터베이스를 이용해 두 그룹의 데이터 세트를 생성했다. Interpatient 데이터 세트는 학습 데이터와 테스트 데이터를 나눌 때 같은 환자의 데이터가 들어가지 않도록 분배하고, inpatient 데이터세트를 나눌 때는 이를 구분하지 않고 분배했다. 두 패러다임에서 심장의 한 주기를 나타내는 비트(beat)는 라벨에 따라서 AFIB 과 AFIB이 아닌 부분으로 라벨링 하였다. 전체 데이터에서 AF은 36%이고, 나머지는 64%이다. 두 패러다임 모두 학습과 테스트를 8:2의 비율로 나누었다. train_test_split이라는 라이브러리에서 random_state와 나누는 비율을 설정하여 진행하였다. 모델 학습 중에는 train 데이터의 20%를 validation 데이터로 사용하여 검증하였다.

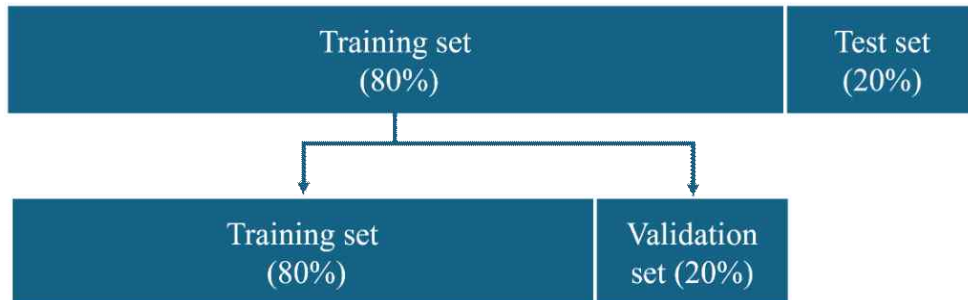


그림18. 데이터 분배 비율

①딥러닝 모델

Model	#Layers	#Parameters	Storage(KB)
Resnet	15	21,698	304
	21	11,138	147
	25	26,978	336

표5. 학습에 쓰인 모델

3가지 모델은 모두 이중분류 학습을 위해 설계되었고, 마지막 층은 동일하다. 손실함수는 binary_cross_entropy가 쓰였고, 마지막 층인 dense layer에는 활성화 함수로 softmax 쓰였고, 최적화 함수로는 Adam이 사

용되었다. 이때 learning rate는 0.0001, batch size는 128, epoch은 100이다. 모델의 입력 데이터의 크기는 (300,1)이고, 출력 데이터는 (2)이다. 출력데이터는 라벨과 동일한 크기이다. label은 0은 정상이고, 1은 심방세동이다. 각 모델의 공통점은 마지막에 global average pooling layer를 통과해준 후, dense layer에서 활성화함수인 softmax를 통해 최종 분류한다. 레이어 15 모델과 다른 두 모델의 차이점은 레이어 15는 잔차연결구조가 한 번있고, 다른 모델은 두 번 있고 global average layer 앞에 max pooling layer가 추가 되었다. (그림 20, 그림 21, 그림 22) 그리고 3 개의 모델은 ResNet 기반으로 잔차 연결구조로 이루어졌다. 잔차 구조는 기존에 학습한 정보를 보존하면서 새로운 정보를 추가로 학습하는 시스템이므로 동일하거나 높은 수준의 정확도를 보인다. (그림19)

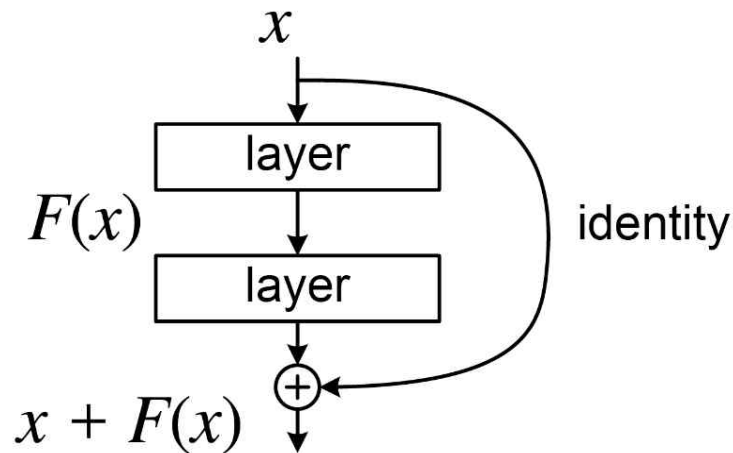


그림19. skip connection

모델은 텐서플로우(Tensorflow) 프레임워크 기반으로 생성하였다. 텐서플로우는 구글이 제공해주는 오픈소스 소프트웨어 라이브러리로, 머신러닝과 딥러닝을 쉽게 사용할 수 있게 한다. 모델 학습에 쓰인 컴퓨터의 사양은 3.69GHz에서 Intel Core i5 -12600KF CPU 이고, 32GB의 RAM을 가진다. GPU는 cuda 11.2가 설치된 NVIDIA RTX3060Ti이다. 학습한 모델은 엡지 디바이스인 젯슨 오린 나노 625MHz에서 6코어 Arm Cortex-A78AE v8.2 64비트 CPU 1.5 MB L2 + 4 MB L3이고, GPU는 32개의 Tensor 코어가 있는 NVIDIA Ampere 아키텍처이다. 젯슨 오리노의 스펙은 표6에 자세히 나와있다.

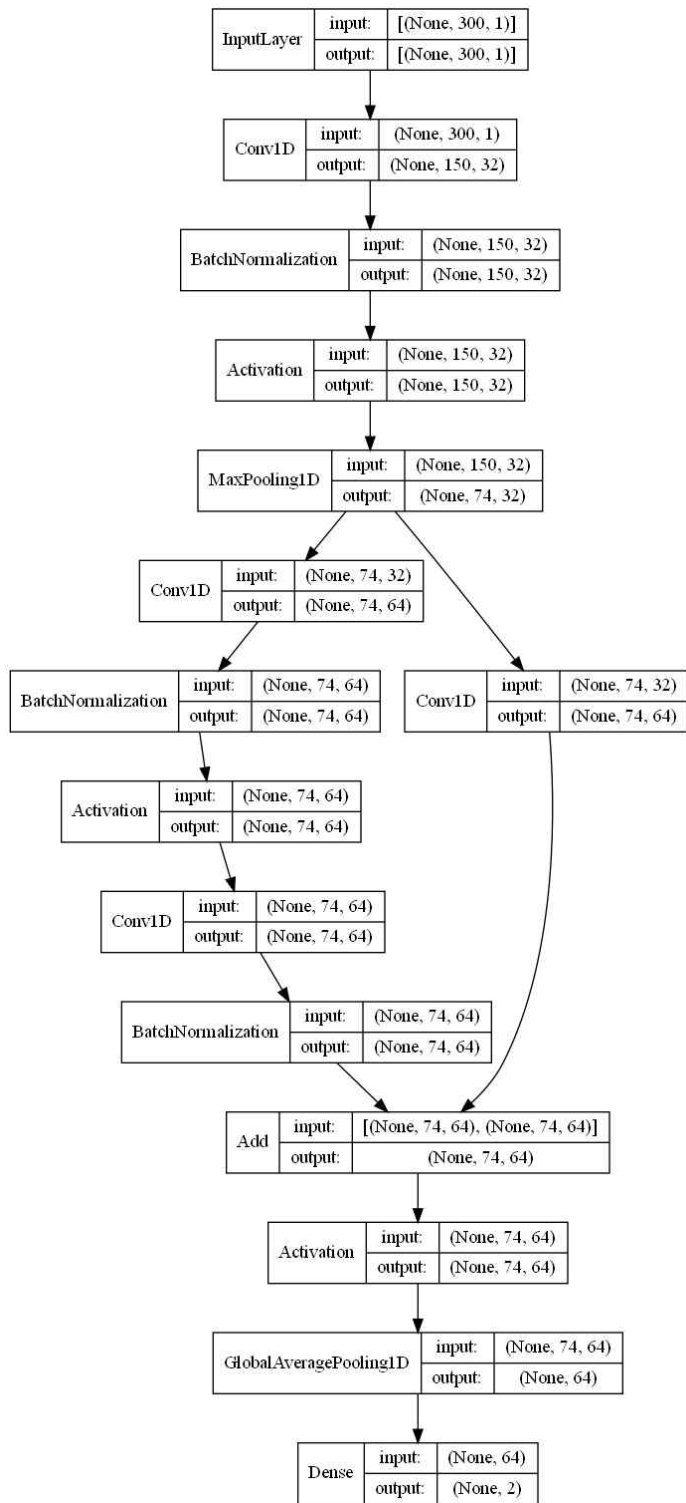


그림20. 15개 레이어 모델의 구조

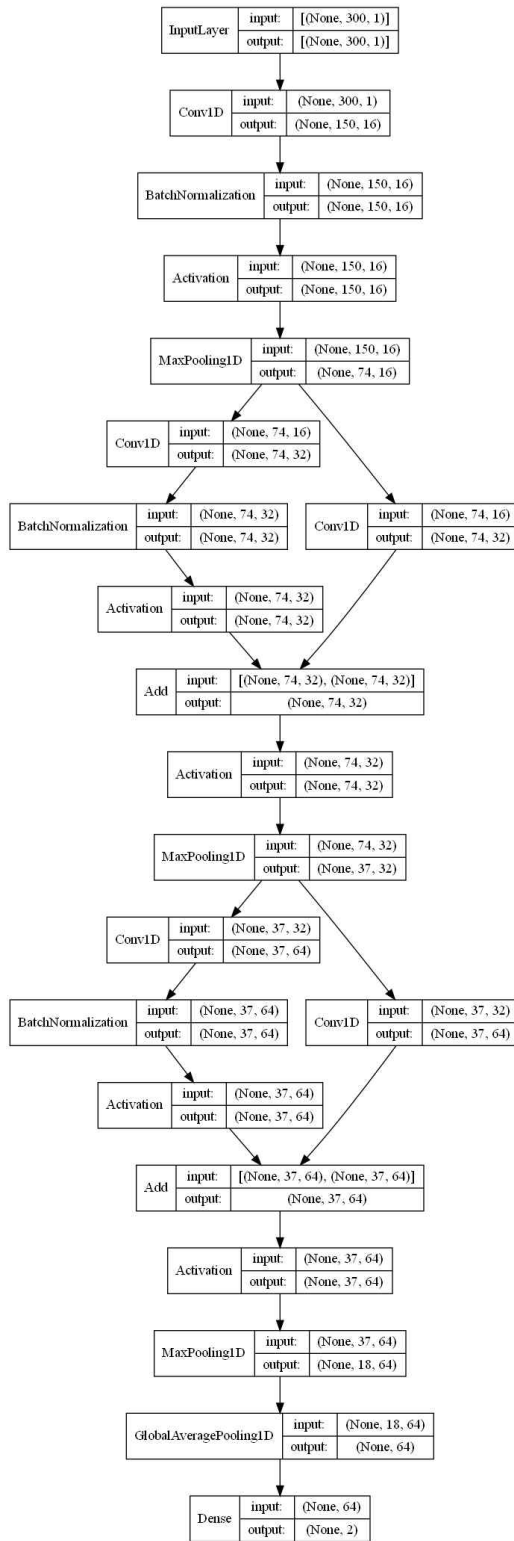


그림21. 21개 레이어 모델의 구조

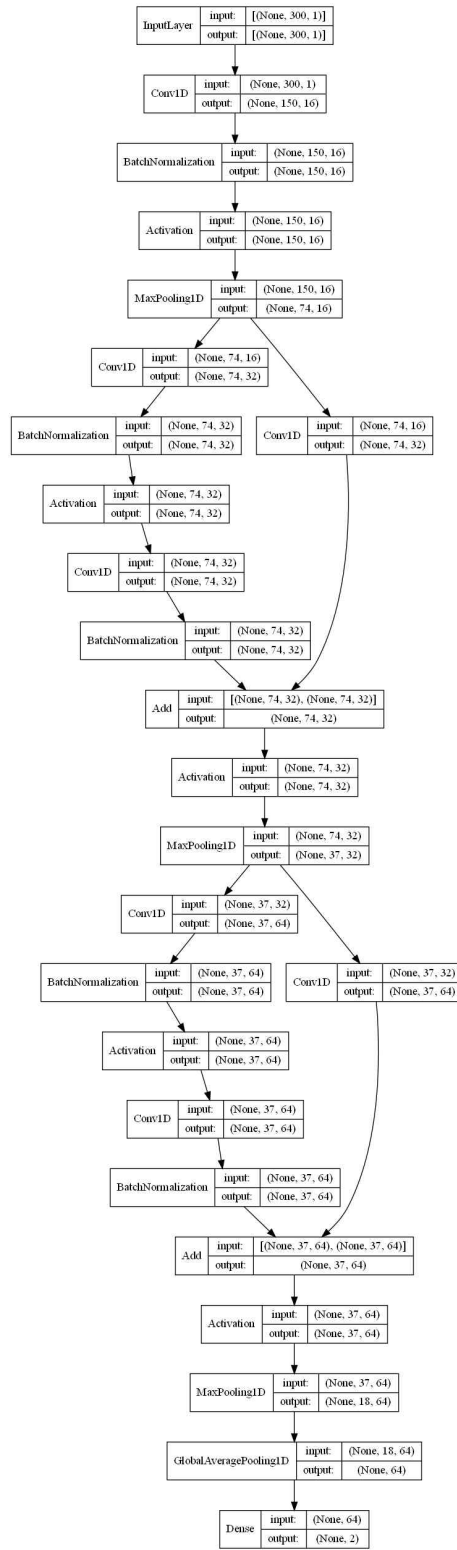


그림22. 25개 레이어 모델의 구조

② 학습 진행

이 실험은 두 개의 패러다임으로 진행했다. Interpatient 와 Inpatient 패러다임이다. Interpatient는 다른 환자들 간에 비교하는 것이고, Inpatient는 환자 내에서 여러 조건 또는 시간에 따라 비교하는 것이다. interpatient는 레코드 아이디에 따라서 분류하였다. 각 패러다임마다 따로 학습을 진행하여, 총 6개의 모델을 학습시켰다. 학습은 외부 장치를 가지고 진행하고, 모델의 가중치값만 keras h5 형식으로 저장하였다. 하드웨어 장치인 젯슨 나노 오린에서 모델과 모델의 가중치를 불러와서 실행시키는 방식으로 test 결과를 확인하였다.

③ 젯슨 나노 오린(Jetson Nano Orin)에서 실행

NVIDIA사의 젯슨 디바이스 중 두번째로 작은 제품이고 이미지 분류, 객체 감지, 분할 및 음성처리와 같은 앱을 위해 신경망 여러 개를 병렬로 실행할 수 있는 작고 강력한 임베디드 컴퓨터이다. 즉, 초소형, 고효율, 저비용으로 AI 구현할 수 있게 해준다. 젯슨 오린 나노의 스펙은 표4에 나타나 있다.



그림23. 젯슨 오린 나노 하드웨어 장치

AI Performance	40 TOPS
GPU	1024-core NVIDIA Ampere architecture GPU with 32 Tensor Cores
GPU Max Frequency	625 MHz
CPU	6-core Arm Cortex-A78AE v8.2 64-bit CPU 1.5MB L2 + 4MB L3
CPU Max Frequency	1.5 GHz
Memory	8GB 128-bit LPDDR5, 68 GB/s
Storage	- (SD Card Slot & external NVMe via M.2 Key M)
Video Encode	1080p30 supported by 1-2 CPU cores
Video Decode	1x 4K60 (H.265) / 2x 4K30 (H.265) 5x 1080p60 (H.265) / 11x 1080p30 (H.265)
CSI Camera	2x MIPI CSI-2 22-pin Camera Connectors
PCIe*	M.2 Key M slot with x4 PCIe Gen3 M.2 Key M slot with x2 PCIe Gen3 M.2 Key E slot
USB*	USB Type-A Connector: 4x USB 3.2 Gen2 USB Type-C Connector for UFP
Networking*	1xGbE Connector
Display	1x DisplayPort 1.2 (+MST) connector
Other I/O	40-Pin Expansion Header(UART, SPI, I2S, I2C, GPIO) 12-pin button header / 4-pin fan header microSD Slot / DC power jack
Power	7W - 15W
Mechanical	100 mm x 79 mm x 21 mm (Height includes feet, carrier board, module, and thermal solution)

표4. 젯슨 오린 나노 하드웨어의 스펙

(3) 연구평가 및 결과

		POSITIVE	NEGATIVE
ACTUAL VALUES	POSITIVE	TP	FN
	NEGATIVE	FP	TN

그림 24. confusion matrix (혼동행렬)

(출처: <https://medium.com/@mudgalvivek2911/machine-learning-confusion-matrix-error-matrix-c518bca18de7>)

학습한 모델의 성능을 평가하기 위해서 비트(beat) 단위의 테스트 데이터를 학습한 모델에 넣었을 때 얼마나 잘 맞추는지를 평가행렬을 통하여 확인 했다. 평가행렬 (Evaluation Matrix)에는 정밀도(Precision), 민감도 (Recall), 특이도(specificity), F1 점수 (F1 -Score), AUC 곡선 (Area Under Curve)을 사용했다.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

식1. 평가지표 계산 공식

심방세동과 심방이 아닌 것으로 이진 분류를 진행한 모델을 평가할 때 정확도만 사용하지 않은 까닭은 정확도 하나만 이용해서 평가할 경우 숫자의 눈높음에 속을 수 있기 때문이다. 예를 들어서 두 개의 클래스를 분류하는 모델이 사용하는 데이터에서 positive 라벨이 하나 일 경우, 모델이 모든 데이터를 negative 로 예측하면, 정확도는 99%가 나올것이다. 이러한 오류를 범하지 않기 위해 정밀도와 재현율을 평가 지표로 이용합니다. 정밀도는 positive를 놓치는 실수를 줄이는데 목표, 얼마나 정밀하게

positive 예측했는지 평가한다. 예를 들어서 데이터의 99%가 정상이어서, 모델이 다 정상이라고 예측할 경우 정밀도는 0이 나온다. 하지만 정밀도만 평가할 경우 negative label에 대한 평가를 내릴 수 없다. 그래서 재현율이 필요한 것이다. 재현율은 positive 라벨의 최대 개수를 찾는 것이 목표고, 문제에서 라벨을 얼마나 제대로 찾았는지 확인한다. 이를 통해 정밀도의 한계를 극복해준다. 이러한 정밀도와 재현율은 서로 보완해주는 관계이므로, 이 둘을 한번에 볼 수 있도록 도와주는 것이 F1- 점수이다. F1- 점수는 정밀도와 재현율의 조화평균이고, 정밀도와 재현율이 최대화 될 때 높은 값이 나온다.

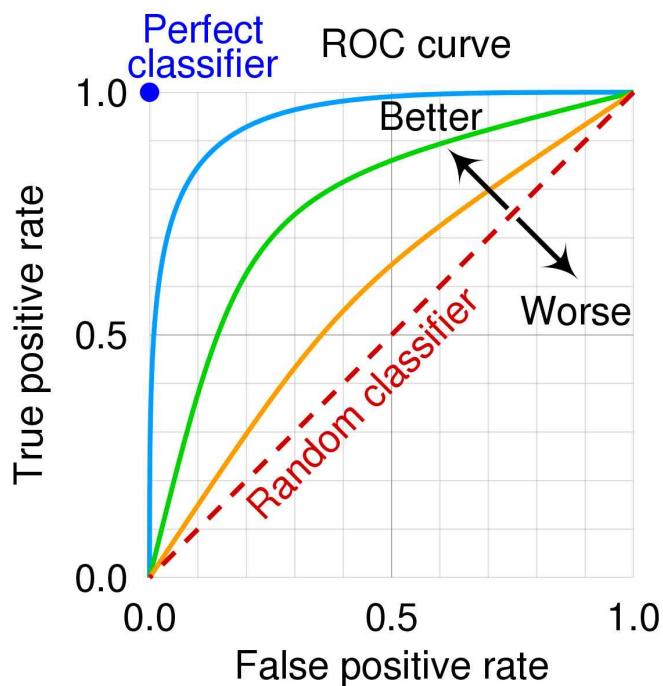


그림25. ROC curve

(출처:https://en.wikipedia.org/wiki/Receiver_operating_characteristic)

AUC 곡선도 평가 지표로 사용하였다. 이 곡선에서 x축은 FPR(False Positive Rate), y축은 민감도 (TPR : True Positive Rate)이다. FPR은 1 - 특이성 (TNR: True Negative Rate)을 의미합니다. ROC 곡선은 가운데 직선에 가까울수록 성능이 떨어지고, 멀어질수록 성능이 높아집니다. 이 좌표를 통해서 보고자 하는 것을 True를 얼마나 True라고 하는지와, False를 얼마나 False라고 잘 판별하는지 확인하는 것이다.

연구 결과

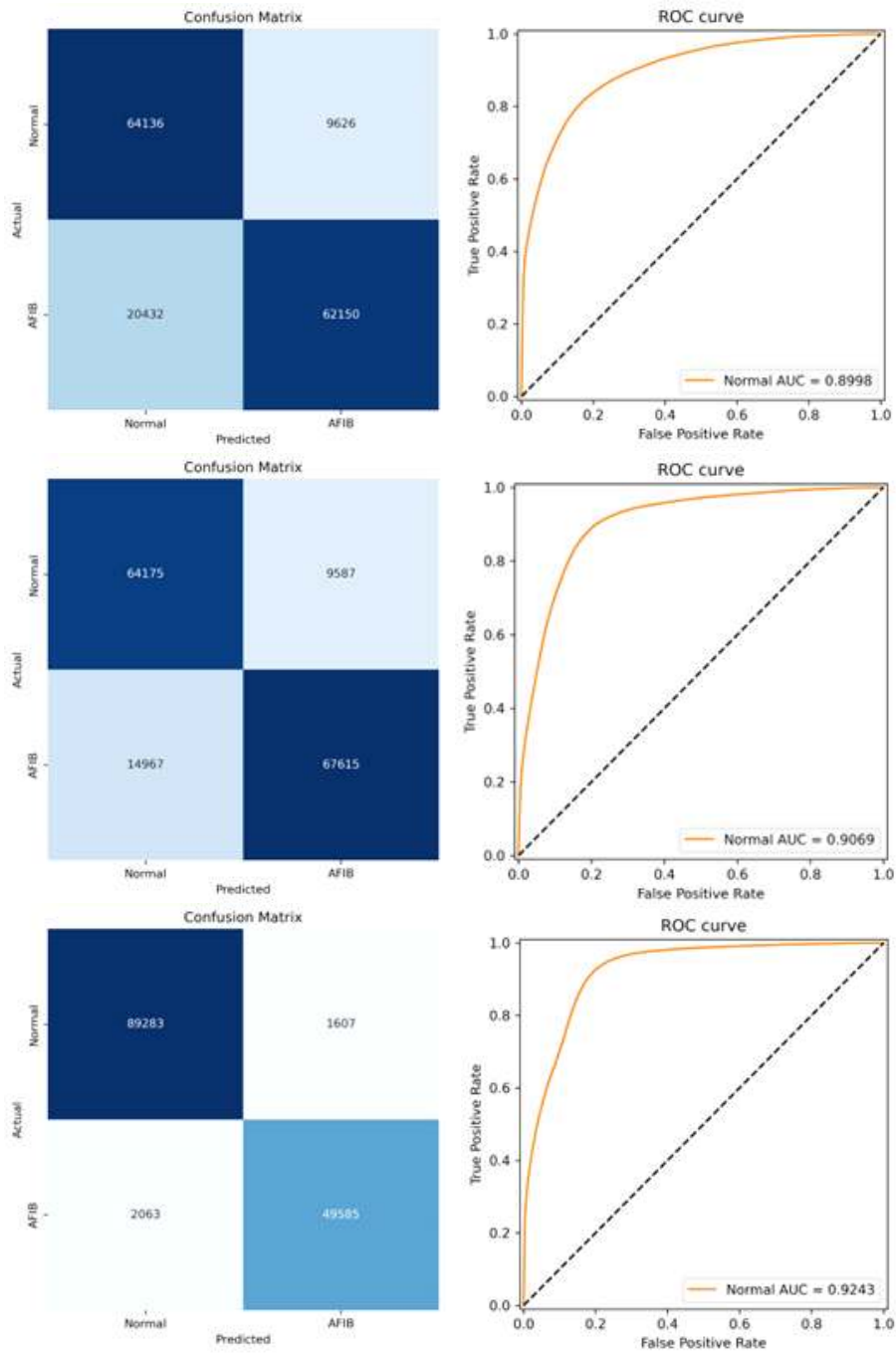


그림26. Interpatients 데이터 세트로 학습한 모델별 결과.(레이어 15, 레이어 21, 레이어 25)

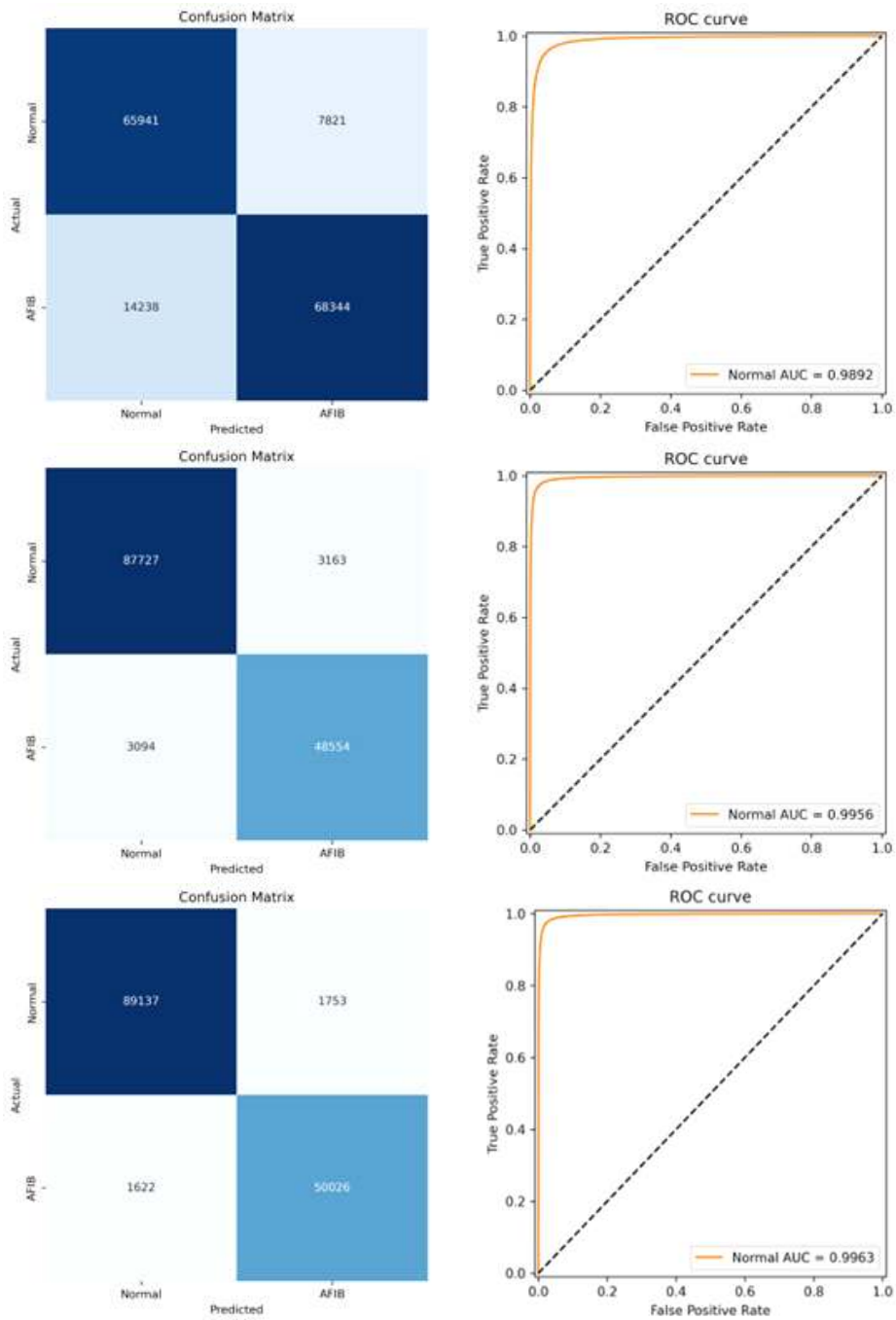


그림27. Inpatients 데이터 세트로 학습한 모델 별 결과 (레이어 15, 레이어 21, 레이어 25)

		Precision	Recall	F1 score
Interpatient	Layer 15	0.8121	0.8110	0.8077
	Layer 21	0.8434	0.8444	0.8429
	Layer 25	0.8599	0.8608	0.8589
Inpatient	Layer 15	0.9524	0.9526	0.9525
	Layer 21	0.9652	0.9651	0.9652
	Layer 25	0.9741	0.9763	0.9763

표7. 3가지 모델을 통해 평가한 정밀도, 재현율, F1 - 점수

토의

이 논문에서는 심전도 데이터를 이용한 심방세동을 가지고, 지도학습 방식으로 경량화 분류 모델을 학습시켰다. 학습은 두 가지 패러다임으로 진행하였는데, Interpatient 데이터 세트와 Inpatient 데이터세트로 나누어서 실험을 진행하였다. 이렇게 두 가지 패러다임을 이용한 까닭은 두 가지의 상황에서 경량화 모델이 어떻게 작동하는지 확인하기 위함이다.

총 3개의 경량화 모델을 이용한 학습 결과는 표7에서도 볼 수 있듯이, Interpatient 데이터 세트로 학습한 결과보다 Inpatient 데이터세트로 진행한 학습 결과의 성능이 높은 것을 알 수 있다. Inpatient로 학습한 결과는 환자 개별 특이성을 무시하였으므로 주로 정확도를 선보인다. 그러나 임상에 적용하기 위해서는 학습과 테스트에 같은 환자가 포함되지 않은 Interpatient 패러다임이 유의미하다.

레이어와 파라미터는 평가 척도 값들과 비례하는 것을 관찰할 수 있다. 하지만 레이어와 파라미터가 증가할 때 연산량과 학습시간과 메모리 용량 또한 증가한다. 그러므로 경량화 모델을 이용하여 하드웨어에 적용할 경우, 최종적으로 설치할 하드웨어의 조건을 고려하여 조절해야 한다.

따라서 연산량과 메모리가 줄어든 경량 모델이 제한된 환경인 하드웨어에서 높은 성능을 나타낸 것을 통해 하드웨어의 제약을 가지는 다른 웨어러블 디바이스에도 적용할 수 있을 것이다.

연구의 한계 및 고찰

이 연구에서는 한 종류의 데이터셋만 사용하였다. 전체 데이터의 양은 많으나 23개의 레코드라는 제한된 수의 환자 정보만 담고 있다. 또한 학습과 평가에 같은 데이터셋에서 획득한 환자만 사용되었기에, 다른 데이터셋을 이용한 평가가 이루어지지 못했다. 뿐만아니라, 데이터셋에서 인종이나 연령대 및 성별에 대한 정보를 제공하지 않기 때문에, 이에 대한 평가도 이뤄질 수 없었다. 이러한 표본 수의 제한성으로 다양성이 부족하다.

게다가 한 종류의 하드웨어로만 평가를 진행하였기에 다른 하드웨어에서의 평가를 통해 경량화 모델의 성능을 비교할 수 있을 것이다. 또한, 본 연구에서는 경량화 모델을 실행시키는 시간과 전력을 명시하지 않았는데, 이는 모델을 구동할 때마다 걸리는 시간과 소모된 전력이 변동하여 정밀한 평가가 어려웠고 평균값을 산출할 수밖에 없었다. 그리고 입력 데이터의 개수에도 영향을 받기에, 시간과 소모 전력을 표시하는 것이 유의미하지 않다고 판단하였다.

이외에도 본 연구에서는 경량 알고리즘 하나인 ResNet 방식만 채택하여 사용했기에 다른 경량 알고리즘 방식과 비교하거나 병합한다면 더 좋은 성능을 낼 수 있을 것으로 기대된다.

결론

종합적 본 연구에서는 두 개의 패러다임인 Interpatient와 Inpatient로 진행하였다. 연구에서 홀터 심전도 데이터를 이용하여 경량화 모델을 학습시키고 심방세동 분류하였다. 학습시킨 모델을 온디바이스인 젯슨 오리노에서 이식하여 평가하였다. 총 3개의 모델의 평가 결과 중 최댓값은 Interpatient 패러다임에서는 F1- 점수가 0.92가 나왔고, Inpatient에서는 0.86이었다. 연산량과 메모리의 차지 비율이 높았던 거대한 모델을 경량화시켜 하드웨어에 성공적으로 이식하는 것을 이외에 다른 웨어러블 디바이스에도 적용할 가능성을 확인하였다.

따라서 심장의 이상을 조기에 진단하여 심장 질환의 발병 위험을 낮추고, 개선하여 생존율을 높일 수 있을 것이다. 게다가 경량화 모델을 이용하여 온디바이스 형태로 환자에 제공할 때 부정맥을 식별하는 각 단계에서 편리성이 증진될 것이다.

참고문헌

- [1] Roth, Gregory A., et al. "Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017." *The lancet* 392.10159 (2018): 1736-1788.
- [2] 통계청 2022년 사망원인 통계결과 [online] https://kostat.go.kr/board.es?mid=a10301060200&bid=218&act=view&list_no=427216
- [3] Sörnmo, Leif, and Pablo Laguna. *Bioelectrical signal processing in cardiac and neurological applications*. Vol. 8. Academic press, 2005.
- [4] "Electrocardiogram." Texas Heart Institute, www.texasheart.org/heart-health/heart-information-center/topics/electrocardiogram/. Accessed 28 May 2024.
- [5] Petmezas, Georgios, et al. "Automated atrial fibrillation detection using a hybrid CNN-LSTM network on imbalanced ECG datasets." *Biomedical Signal Processing and Control* 63 (2021): 102194.
- [6] Hughes JW, Olgin JE, Avram R, et al. Performance of a Convolutional Neural Network and Explainability Technique for 12-Lead Electrocardiogram Interpretation. *JAMA Cardiol.* 2021;6(11):1285-1295. doi:10.1001/jamacardio.2021.2746

- [7] Du, Mingyu, Yuan Yang, and Lin Zhang. "Diagnosis of atrial fibrillation based on unsupervised domain adaptation." *Computers in Biology and Medicine* 164 (2023): 107275.
- [8] Fatima, Rumsha, et al. "Advancing ECG Diagnosis Using Reinforcement Learning on Global Waveform Variations Related to P Wave and PR Interval." *arXiv preprint arXiv:2401.04938* (2024).
- [9] Naveed, Humza, et al. "A comprehensive overview of large language models." *arXiv preprint arXiv:2307.06435* (2023).
- [10] Zhao, Wayne Xin, et al. "A survey of large language models." *arXiv preprint arXiv:2303.18223* (2023).
- [11] Wang, Ching-Hao, et al. "Lightweight deep learning: An overview." *IEEE consumer electronics magazine* (2022).
- [12] Iandola, Forrest N., et al. "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size." *arXiv preprint arXiv:1602.07360* (2016).
- [13] Han, Kai, et al. "Ghostnet: More features from cheap operations." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.
- [14] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). *PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals*. *Circulation* [Online]. 101 (23), pp. e215-e220.

[15] Mayapur, Priyanka. "A Review on Detection and Performance Analysis on R-R Interval Methods for ECG." (2018).

[16] Haddadi, Rachid, et al. "Discrete wavelet transform based algorithm for recognition of QRS complexes." *2014 International Conference on Multimedia Computing and Systems (ICMCS)*. IEEE, 2014.

영문초록 (Abstract)

Deep learning models have evolved by increasing the number of parameters and layers to improve performance, thereby increasing the complexity of networks. The utilization of advanced deep learning models in the field of medical artificial intelligence has gradually increased. Now, we are witnessing an era where deep learning models are not only operated on computers or in the cloud but also on-device, in hardware form. There is a growing trend among companies to apply on-device models, which can be executed directly on the device, providing fast response times, low power consumption, low cost, and enhanced privacy protection. However, on-device artificial intelligence faces hardware constraints such as power consumption, RAM usage, storage space, and computing power. To overcome the limitations of edge devices, models are lightweighted, enabling faster processing times and handling more processing requests. Heart disease is one of the top three causes of death in our country. Among them, atrial fibrillation, a condition where the atria contract irregularly and rapidly, is a common type of arrhythmia. In this study, we aim to train and evaluate a model using data from patients with atrial fibrillation. We utilize the MIT-BIH Atrial Fibrillation Dataset, an open data format provided by Physionet. To explore both interpatient and inpatient paradigms, we create two training datasets. Then, we perform segmentation into 1-beat segments using RR intervals for each dataset. A binary classification is conducted to differentiate between atrial fibrillation and non-atrial fibrillation using the segmented data. The training is carried out using supervised learning, where the program learns from labeled data to make predictions. We employ three models based on Resnet with skip connections for supervised learning. The trained models are evaluated on the hardware device, Jetson Nano Orin, after setting hyperparameters: learning rate of 0.0001, batch size of 128, binary cross-entropy loss function, and softmax activation

function in the final dense layer. The research results are evaluated using key metrics for classification models: Precision, Recall, F1 score, Accuracy, and AUC. The results from the inpatient dataset, where records from the same patient are mixed, show significantly better performance compared to those from the outpatient dataset. The best performance achieved in the inpatient dataset is an F1 score of 0.9763. In contrast, the best-performing model in the outpatient dataset achieves an F1 score of 0.8589. Through this, it was practically confirmed that the atrial fibrillation classification model can be executed on on-device hardware. Limitations of this study include training with only one type of dataset, evaluation of lightweight models on only one type of hardware, and selecting only one method of lightweighting. Despite these limitations, it is clinically significant as it can reduce the time and cost involved in identifying atrial fibrillation from prolonged electrocardiogram recordings and can execute the model on on-device hardware regardless of location and environment.