

심장 소리 분류를 위한 Inverted Residuals 기반 경량화 모델

박두서¹ · 이민영² · 김기현¹ · 이홍철^{1**}

¹고려대학교 산업경영공학과 / ²카카오브레인

A Lightweight Model for Heart Sound Classification Based on Inverted Residuals

Doo-Seo Park¹ · Min-Young Lee² · Ki-hyun Kim¹ · Hong-Chul Lee¹

¹Department of Industrial and Management Engineering, Korea University

²Kakao Brain

For the treatment and prevention of heart diseases, which is the leading cause of high mortality rate globally, the need for healthcare devices equipped with artificial intelligence(AI) model that can monitor in real-time and analyze heart conditions is increasing. Therefore, in this study, we propose a light CNN that can be applied to healthcare devices, using the PASCAL data. The proposed model used MFCC feature extraction method suitable for heart sound range, The light CNN was designed with the inverted residuals used in MobileNetV2. The experiments showed that the proposed model with fewer 82.5% of the learnable parameters, achieved similar performance in accuracy within the range of 1 to 2% compared to the previous studies. It was confirmed that the proposed light CNN can be feasibly incorporated on mobile devices by means of comparative experiments in a reasonable amount of computation.

Keywords: Heart Sound Classification, Inverted Residuals, Lightweight Model Architecture, PASCAL Classifying Heart Sounds Challenge

1. 서론

국내 사망 원인 2위인 심장 질환은 암 다음으로 높은 순위를 차지하고 있으며, 2009년부터 심장 질환에 의한 사망률은 꾸준히 증가하고 있다. 그리고 2019년 통계에 따르면 급성심장정지 환자의 생존률은 약 8.7%이며 생존한 환자의 뇌기능 회복률은 약 5.4%이다. 뇌기능 회복률이란 급성심장정지 환자가 퇴원 당시 혼자서 일상생활이 가능할 정도로 회복된 상태를 의미한다. 이에 개인의 건강과 의료에 관한 다양한 플랫폼을 다루는 산업 분야와 의료 IT가 융합된 형태인 스마트 헬스케어 통해 심장 건강 상태를 모니터링하고 분석할 수 있는 서

비스가 제공되고 있다(Ha and Lee, 2020). 스마트 헬스케어는 의료 서비스 패러다임을 기존 사후 치료방식에서 의료 빅데이터를 기반으로 예방·건강 관리 중심으로 변화시키고 있다. 특히 모바일 기술의 발달로 인해 점점 소형화 및 경량화되고 있는 헬스케어 디바이스에 빅데이터와 인공지능을 활용하여 기존 의료기기보다 성능 및 효율이 크게 향상되어 진단의 정확성은 높아지고, 질병 예측 및 예방 등 새로운 의료 정보의 가치 창출을 통해 새로운 개념의 의료 서비스 시장을 창출하고 있다(Kwon, 2021).

스마트 헬스케어 산업의 발달에 따라 심장 소리의 이상 유무를 분류하는 연구가 다양하게 시도되고 있다. 하지만, 심장

이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구 (No. 2020-0-02199, 비대면 진료환경 구축을 위한 AI기반 심폐음 질병분류 모델 개발)이며, 본 연구는 4단계 두뇌한국21에 의해 지원되었습니다.

* 연락저자 : 이홍철 교수, 02841 서울특별시 성북구 안암로 145 고려대학교 산업경영공학부, Tel : 02-3290-3389, Fax : 02-929-5388,

E-mail : hcleee@korea.ac.kr

2021년 9월 14일 접수; 2021년 10월 25일 수정본 접수; 2021년 11월 18일 게재 확정.

소리 이상 유무를 분류하는 기존 연구들은 크게 2가지의 한계점이 있다. 먼저, 대다수의 심장 소리 이상 유무를 분류하는 연구는 심장 소리를 분할하여 분할된 주기에서 추출한 통계치를 특징으로 사용하고 있다. 심장 소리를 분할하기 위해 심전도(ECG), 엔벨로프(Envelope) 기반 분석, 통계적 모델을 주로 활용하고 있다. 하지만 심전도를 측정하기 위해서는 전문적인 의료장비가 필요하며 일반인이 스스로 측정하기에는 제한사항이 많다. 엔벨로프 기반 분석 방법은 주변 소음 및 비정상 심장 소리에서 발생하는 잡음 등에 영향을 받으며, 일반 성인과 심장 소리의 패턴이 다른 영유아 또는 신생아, 심장병 환자에게 잘 적용되지 않는다. 마지막으로 통계적 모델을 활용하는 방법은 Fundamental Heart Sound(FHS)의 특성이 다양하여 모든 FHS의 모델링에 한계가 있다(Deng and Han, 2016). 그리고 비정상 심장 소리 중 일부는 심장 주기마다 발생하지 않고 간헐적으로 발생하는 것으로 알려져 있다(Bentley *et al.*, 2011). 이에 심장 소리를 분할하지 않고 이상 유무를 분류할 수 있어야 한다.

두 번째로, 헬스케어 디바이스의 소형화 및 경량화를 고려한 연구가 미비하다. 기존의 심장 소리를 분류하는 연구는 주로 머신러닝 기법을 사용하고 있으며, 딥러닝을 제안한 연구 또한 성능에 관한 연구결과만 제시할 뿐 경량화를 고려하지 않는다. 이에 최근 헬스케어 디바이스에 탑재할 수 있는 기존 모델보다 경량화되며, 분류 성능은 유지 가능한 모델에 대한 연구가 시도되고 있지만, 파라미터 수와 연산량만 반영할 뿐, 모바일 기기에 탑재하기 위한 모델 파일 크기 및 추론 시간 등 실질적인 경량화에 대한 평가에 한계가 있다. 이에 본 연구에서는 기존의 경량화 연구에서 사용한 지표인 학습 파라미터 개수와 연산량(MAdds)과 모델 파일 크기, 추론 시간 및 모델 정확도를 포함하여 총 5가지 지표를 통해 모델을 정량적으로 평가하였다.

따라서, 본 연구에서는 심장 소리를 분할하지 않고 이상 유무를 분류할 수 있으며, 헬스케어 디바이스에 탑재가 가능한 경량화된 딥러닝 모델을 제안한다.

본 논문의 구성은 다음과 같다. 제2장에서는 심장 소리 분류와 관련 연구에 대해 설명하고, 제3장에서는 데이터 소개, 데이터 전처리 및 특징 추출과 모바일 넷 V2(Sandler *et al.*, 2018)에서 사용한 Inverted Residual Block을 적용하여 CNN 모델을 경량화한 방법론에 대해 소개한다. 실험 결과는 제4장에서 서술하며, 마지막으로 제 5장에서는 본 연구의 결론을 다루도록 한다.

2. 관련 연구

본 연구는 헬스케어 디바이스에 탑재가 가능한 경량화된 딥러닝 모델을 제안하고자 한다. 이에 2012년에 열린 PASCAL Classifying Heart Sounds Challenge에서 공개된 데이터를 사용하

였다. 대회 당시 Precision, Total Precision, Specificity, Sensitivity, Youden's Index, Discriminant Power 등을 평가지표로 결과를 제시하였다. 본 장에서 설명할 선행연구들의 평가지표는 Total Precision을 기준으로 평가하였다. 평가지표에 대한 자세한 설명은 제 4장 실험 및 실험 결과에서 서술하였으며, 본 장에서는 PASCAL 데이터를 사용하여 연구된 머신러닝 및 딥러닝 알고리즘 기반 방법론에 대해 설명한다.

2.1 PASCAL 데이터를 사용한 머신러닝 모델 연구

머신러닝 연구에서 사용하는 주요 방법론은 심장 소리의 제 1심음(S1)과 제 2심음(S2) 주기를 추출하여 추출된 주기들의 통계치들을 특징으로 활용한다.

Gomes는 각 심장 소리의 S1과 S2 주기를 추출한 이후 총 6개의 특징(S1 · S2 주기들의 평균, S1 · S2 주기들의 표준편차, 중앙값, R^2)을 이용하여 의사결정나무(Decision Tree)와 다층 퍼셉트론을 통해 심장 소리를 분류하는 방법론을 제안하였다. 의사결정나무에 비해 다층 퍼셉트론이 더 우수한 성능을 보였으며, 그 결과 Total Precision이 Data set A에서 2.12, Data set B에서 1.67의 성능을 확인하였다(Gomes and Pereira, 2012). Deng은 심장 소리 분류를 위하여 S1, S2 주기를 검출한 이후 주기의 개수, S1과 S2 사이의 길이와 같은 통계적 특징에 따라 심장 소리를 분류하는 순서도를 제안하였다. Deng이 제안한 방법론의 경우 Total Precision 기준 Data set A에서 1.46, Data set B에서 1.31의 분류 성능을 보여주었다(Deng and Bentley, 2012).

심장 소리를 S1과 S2 주기로 분할한 Balili는 S1, 수축기, S2, 확장기의 4구간에 대한 이산형 웨이블릿 변환을 통해 구한 100가지의 특징들을 랜덤 포레스트(Random Forest)를 이용하여 분류하였으며, Total Precision을 기준으로 Data set A에서는 2.92를, B에서는 2.04의 성능을 보였다(Balili *et al.*, 2015). 심장 소리의 신호를 6차 Daubechies 웨이블릿 변환을 통해 재구성한 이후 Average magnitude difference function(AMDF)을 통해 S1, S2 주기를 추정하고 S1을 기준으로 분할된 심장 주기를 정렬시킨다. 그 이후 서로 다른 길이의 심장 주기에서 추출된 스펙트로그램(Spectrogram)의 크기를 쌍선형 보간법을 통해 동일하게 맞춰주고 Partial least squares regression(PLSR)을 통해 차원을 축소시켰다. 분류 모델은 Support Vector Machine(SVM)을 사용하였고, Total Precision 기준 Data set A에서 2.89를, B에서는 1.75의 분류 성능을 보였다(Zhang *et al.*, 2017). Zhang은 심장 소리를 Band Pass Filter를 통해 잡음을 제거하고 스펙트로그램을 추출하여 쌍선형 보간법을 통해 서로 다른 크기의 스펙트로그램을 같은 크기로 맞춰준 이후 텐서 분해를 통해 차원을 축소하고 SVM 분류기를 사용하는 방법론을 제안하였다. 분류 결과는 Total Precision 기준으로 Data set A에서 2.9를, B에서 1.68의 성능을 보였다(Zhang *et al.*, 2017a).

Chakir은 데이터 세트 A와 B의 특성에 맞는 심장 소리를 분할하는 알고리즘을 제안하였고, 이를 통해 분할된 심장 소리에서 S1과 S2 사이의 길이, zero crossing, 각 데이터의 길이 및 진폭의 최소, 최대값 등의 특징을 추출하여 K-Nearest Neighbor(KNN) 기법을 통해 분류하였다. Total Precision을 기준으로 Data set A에서 2.96과 B에서 1.58의 결과를 보였다(Chakir *et al.*, 2018). Bourohou는 4가지 엔벨로프 방법을 통해 심장 소리를 분할하였으며, 분할된 심장 소리에서 20개의 특징을 추출하였다. 분류 모델은 KNN, SVM, 트리 기반 분류기, 나이브 베이즈 분류 등 총 4가지의 모델을 제안하였다. 나이브 베이즈 분류기에서 가장 좋은 결과를 얻었으며, Total Precision 기준 Data set A에서 3.06과 B에서 2.37로 선행연구들과 비교해 가장 우수한 성능을 보여주었다(Bourohou *et al.*, 2020).

2.2 PASCAL 데이터를 사용한 딥러닝 모델 연구

Zhang은 Band Pass Filter를 적용하여 심장 소리의 잡음을 제거하고 심장 주기를 검출한 이후 단시간 푸리에 변환(Short Time Fourier Transform, STFT)을 통해 스펙트로그램을 추출하였다. 이후 Convolution Layer와 Fully Connected Layer를 결합한 모델을 제안하였으며 Total Precision 기준 Data set A에서 2.96과 B에서 1.62의 성능을 확인하였다. 제 4장 실험 및 실험 결과에서 Zhang의 연구와 본 연구에서 제안한 Inverted Residuals 기반 경량화 방법론과 비교하였다(Zhang and Han, 2017b).

Vrbancic는 Data set A의 심장 소리를 150ms 길이로 분할 후 1D-CNN과 Fully Connected Layer를 결합한 모델을 통해 자동으로 심장 주기를 탐지하는 방법론을 제안하였다(Vrbancic and Podgorelec, 2019). Raza는 라벨링 된 Data set B를 사용하여 4000Hz로 리샘플링하고 Band Pass Filter를 이용하여 잡음을 제거하였다. 서로 다른 데이터의 길이를 짧은 데이터는 반복시키고 긴 데이터는 해당 길이까지만 사용하는 Repeat 기법을 적용하여 모든 데이터를 동일한 길이로 만들었으며, 높은 차원인 데이터를 낮은 샘플링을 통해 차원을 축소하고 RNN을 사용하여 분류하는 방법론을 제안하였다. 분류 성능은 정확도로 측정하여 약 80.8%의 성능을 보였다(Raza *et al.*, 2019).

2.3 경량화 관련 심장 소리 분류 연구

Xiao는 심장 소리 데이터를 2000Hz로 리샘플링하고 잡음을 제거하기 위해 Band Pass Filter를 통과시켰다. 그리고 3초 길이로 분할한 1차원의 PCG 데이터를 입력 데이터로 사용하였다. 모델을 경량화하기 위해 모바일 넷 V2에서 사용한 Inverted Residual 구조에서 Skip Connection이 아닌 Clique block을 적용

하여 모델이 양방향으로 Feature map을 학습할 수 있도록 하였다. Majority voting 방식을 이용하여 최종 클래스를 분류하였다. 적은 파라미터를 학습하면서, 성능의 우수성을 보여주었다(Xiao *et al.*, 2020). Shouvo는 다양한 출처를 통해 심장 소리 데이터를 수집하였다. 데이터는 정상 소리와 Aortic stenosis(AS), Mitral regurgitation(MR), Mitral stenosis(MS), Mitral valve prolapse(MVP)의 4종류의 비정상 소리로 구성하였다. 데이터는 2000Hz로 리샘플링되며, 모두 같은 길이로 맞춰주었다. 분류 모델은 심장 소리의 특징을 추출하는 병렬 CNN 구조와 추출된 특징을 분류하는 Bi-LSTM을 제안하였다. 그리고 PhysioNet 데이터와 Github PCG 데이터를 통해 모델의 우수성을 검증하였고, 0.67M의 파라미터와 26M FLOPS를 학습에 사용하여 웨어러블 기기 및 모바일 기기에 적합한 구조를 가지고 있다(Shuvo *et al.*, 2021). Li는 심장 소리의 잡음을 제거하기 위해 10Hz 이상의 대역대를 갖는 High Pass Filter를 통과시키고 STFT를 적용하여 특징을 추출하였다. 분류 모델은 파라미터의 수를 감소시키기 위해 Convolution Layer를 3개 층만 쌓은 2D-CNN을 사용하였다. 그리고 클래스의 불균형을 보완하기 위해 모델을 학습하기 위한 손실 함수로 FocalLoss 함수를 사용하였다. 선행연구들에 비해 파라미터는 약 90% 감소하였으나 성능은 약 12% 하락하는 것을 확인하였다(Li *et al.*, 2021).

머신러닝과 딥러닝 관련 선행연구들은 전반적으로 성능 향상을 위한 모델 연구를 수행하였으며, 경량화 모델은 모바일 환경을 고려한 연구가 미비하였다. 따라서, 본 연구에서는 모바일 측정기기에 탑재될 수 있는 단순화된 전처리를 적용한 Inverted Residuals 기반 경량화 모델을 제안한다.

3. 제안 방법론

<Figure 1>은 본 연구에서 제안하는 방법론의 전체적인 Framework이다. 심장 소리 분류를 위한 데이터 전처리, 특징 추출, 분류 모델 학습 등을 다음의 과정을 통해 수행한다.

1. 데이터는 PASCAL 데이터를 사용하였으며, Data set A는 Normal, Murmur, Extra Heart Sound, Artifact의 4개의 클래스, Data set B는 Normal, Murmur, Extrasystole의 3개 클래스로 구성된다.
2. 전처리는 Band Pass Filter를 통해 심장 소리의 이상 잡음을 제거하며, 이후 Repeat 기법을 적용하여 모든 데이터를 동일한 길이로 맞춰준다.
3. 특징 추출은 입력된 신호를 특정 주파수 구간의 에너지 계수 형태로 나타내는 MFCC를 사용한다.
4. 분류 모델은 CNN 기반의 Inverted Residual 구조를 사용한 경량화 모델을 통해 심장 소리 이상 유무를 분류한다.

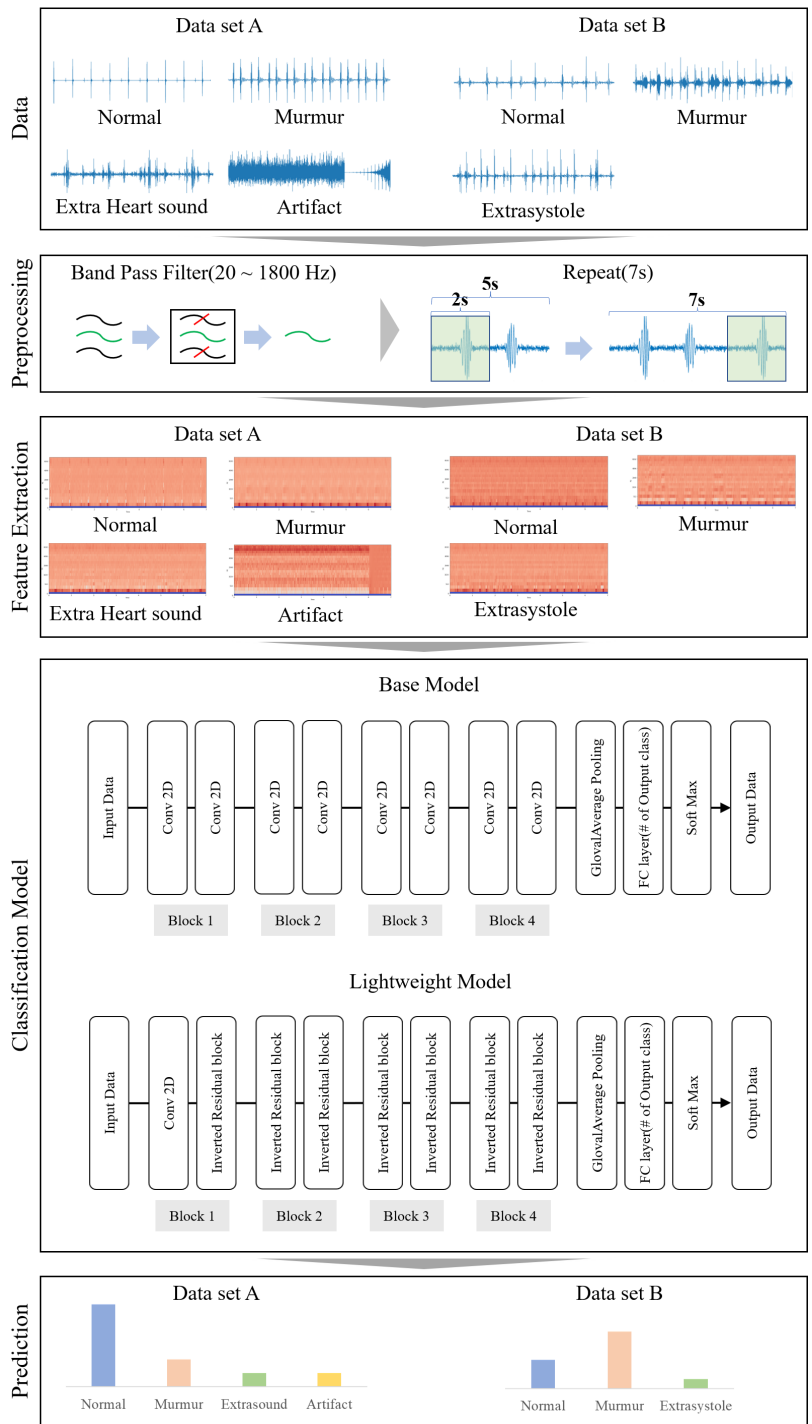


Figure 1. The architecture of the Framework for Heart Sound Classification

3.1 데이터

본 연구는 헬스케어 디바이스에 탑재할 수 있는 모델을 제안하는 것이 주된 목적이므로, 실제 사용자들의 심장 소리 녹음 환경과 유사하게 녹음된 데이터인 PASCAL Classifying Heart Sounds Challenge(PASCAL 심장 소리 데이터)를 이용하였다. 데이터는 Data set A와 B로 구분되어 있으며, Data set A

는 iStethoscope를 활용하여 녹음된 데이터로 Normal, Murmur, Extra Heart Sound, Artifact로 분류되어 있고, Data set B는 Litmann Model 3100 전자 청진기를 활용하여 녹음된 데이터로 Normal, Murmur, Extrasystole로 분류되어 있다. 공통적으로 Data set A와 B는 경연대회 당시 점수 산정용으로 사용한 라벨링 되지 않은 데이터가 존재한다. 이에 라벨링 된 데이터를 학습 데이터와 검증 데이터로 나누었으며, 검증 데이터는 모델

Table 1. The number of samples in the Train, Validation and Test Data set

Data set	Class	Train Data	Validation Data	Test Data (unlabeled data)
Data set-A	Normal	24	7	52
	Murmur	28	6	
	Extra Heart Sound	15	4	
	Artifact	32	8	
Total		99	25	52
Data set-B	Normal	256	64	195
	Murmur	76	19	
	Extrasystole	36	10	
Total		368	93	195

이 학습 데이터에 과적합하는 것을 방지하기 위해 사용하였다. 라벨링 되지 않은 데이터는 평가 데이터로 사용하였다. 전체 데이터의 세부 내용 및 학습 데이터, 검증 데이터 및 평가 데이터 비율은 <Table 1>과 같다(Bentley *et al.*, 2011).

3.2 전처리

심장 소리 데이터를 분류 모델의 입력값으로 사용하기 위하여 전처리를 진행하였다. 먼저 데이터를 4000Hz로 리샘플링한 이후 소리의 잡음을 제거하기 위해 20~1800Hz 대역대를 갖는 Band Pass Filter를 통과시켰다. 그리고 서로 다른 데이터의 길이를 동일하게 맞추기 위하여 전체 데이터의 평균인 7초를 기준

으로 Repeat 기법을 적용하였다. Repeat 기법이란 특정 데이터의 길이가 기준 시간보다 긴 경우 기준 시간 이후 데이터는 잘라내고 짧은 경우에는 기준 시간까지 데이터를 반복하는 기법을 의미한다(Raza *et al.*, 2019). CNN 모델의 입력 데이터로 사용하기 위해서는 모든 데이터를 동일한 길이로 변환하여야 하는데 Repeat 기법은 심장 소리의 주기 검출과 같은 기법을 적용하지 않는 간단한 전처리로 모든 데이터를 동일한 길이로 변경할 수 있다는 장점이 있다. 전처리 적용 전후의 심장 소리 데이터의 도식화는 <Figure 2>와 같다. (a)는 심장 소리가 기준 시간인 7초 이상의 데이터로 전처리 결과 7초 이후는 사용하지 않는 것을 확인할 수 있다. (b)는 7초보다 짧은 경우로, 부족한 시간만큼 데이터를 반복하였다.

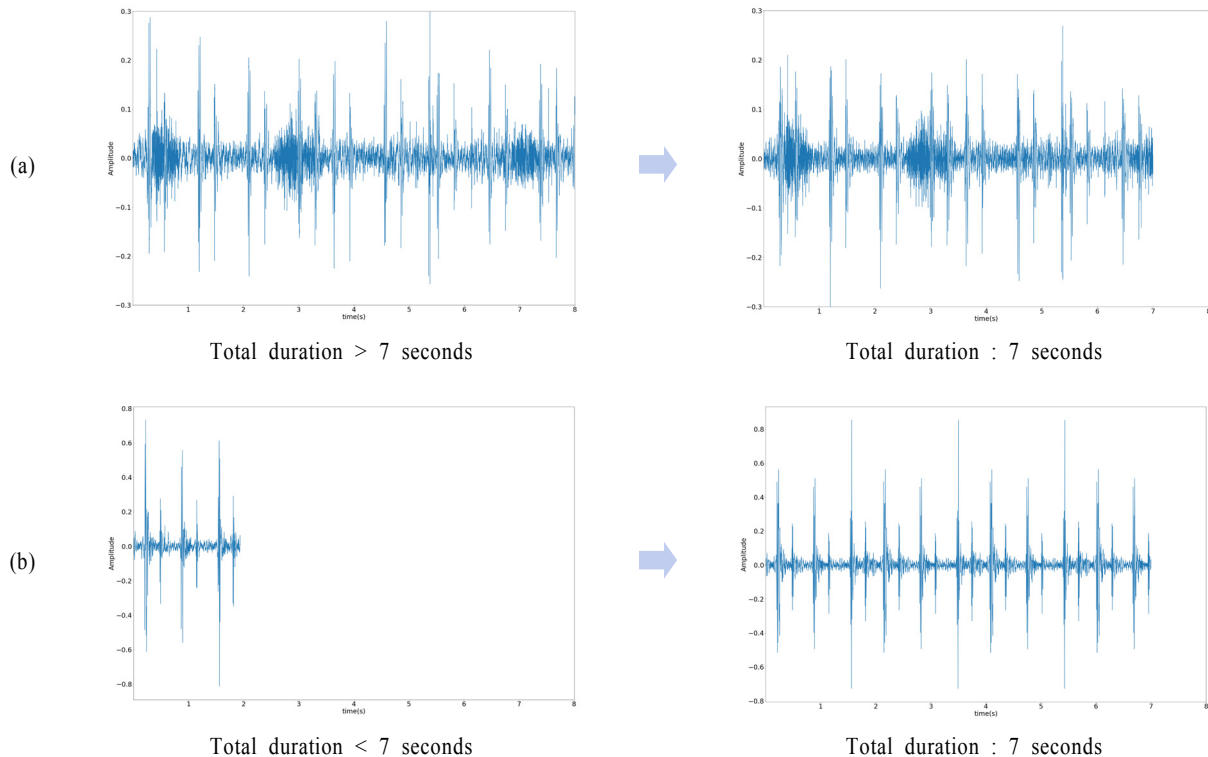


Figure 2. Preprocessing of Heart Sound Signal

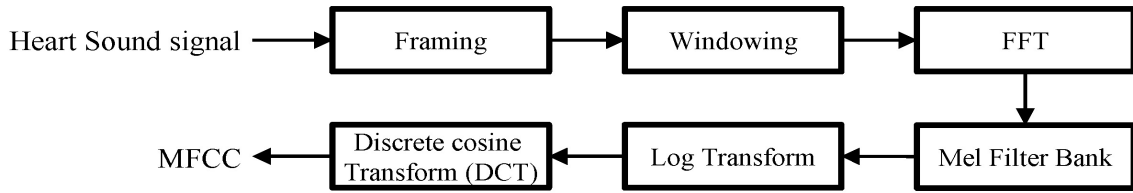


Figure 3. Block Diagram of MFCC Algorithm

3.3 특징 추출(MFCC)

Repeat 기법을 통해 데이터의 길이를 7초로 통일시킨 후 1D 신호를 2D 시간-주파수 신호로 변환시켜주기 위하여 MFCC (Mel-Frequency Cepstral Coefficient)로 특징을 추출하였다. MFCC는 입력된 신호의 단위 에너지를 추출하는 방법으로, 사람이 소리를 인식할 때 저주파수 대비 고주파수의 변화에 덜 민감한 특징을 이용한 Mel Filter Bank를 통해 입력된 신호를 특정 주파수 구간의 에너지 계수 형태로 나타내는 것이다. MFCC는 이산 코사인 변환(Discrete Cosine Transform, DCT)을 통해 압축된 에너지 계수를 특징으로 사용하는데, DCT는 저주파의 에너지를 압축하는 데 강점이 있으므로, 심장 소리 분류에 적합한 특징이며, 에너지를 압축해주기 때문에 차원을 축소해주는 장점이 있다(Kang *et al.*, 2020). MFCC를 추출하는 과정은 <Figure 3>과 같다.

먼저 전처리한 심장 소리 데이터를 짧은 시간 간격으로 분할하는 Framing을 거친다. 이때, 프레임의 길이는 32ms ($0.032 \times 4000 = 128$)이며, 다음 프레임으로 이동 시 이전 프레임과 24ms(96개 샘플)를 겹쳐서 Framing을 한다.

다음은 각 프레임별로 Window 함수와 Element-wise 곱 연산을 수행하는 Windowing 과정을 거친다. 본 연구에서 사용한 Window 함수는 식 (1)의 Hanning Window(Oppenheim, 1999) 함수($w(n)$)이며, N은 프레임 내 샘플의 개수이며, N은 128로 설정하였다.

$$w(n) = 0.5(1 - \cos(\frac{2\pi n}{N-1})), n = 0, 1, \dots, N-1 \quad (1)$$

Window 필터를 통과한 각 프레임의 심장 소리 데이터를 주파수 영역으로 바꾸기 위해 고속 푸리에 변환(Fast Fourier Transform, FFT)을 적용한다. FFT는 이산 푸리에 변환(Discrete Fourier Transform, DFT)을 빠르게 수행할 수 있는 알고리즘이다. FFT를 통해 주파수 영역으로 바뀐 스펙트럼($X(k)$)을 파워 스펙트럼(Power)으로 변환하기 위하여 식 (2)를 적용하여 계산한다. 일반적으로 K는 Framing 과정에서 설정한 프레임의 길이와 동일하게 설정하여, 128의 값을 갖는다(Deng *et al.*, 2020).

$$Power = \frac{|X(k)|^2}{K}, k = 1, 2, \dots, K \quad (2)$$

파워 스펙트럼에 Mel Filter Bank를 통과하여 주파수 범위별 에너지를 추출한다. Mel Filter Bank는 식 (3) ~ (5)를 통해 정의된다. 식 (3)은 주파수(f)를 Mel(M)로 변환하는 수식이다. 식 (3)을 이용하여 0부터 2,000Hz 사이의 주파수 대역을 Mel로 변환한 이후 일정한 간격으로 분할한다. 이후 Mel(M)을 주파수(f)로 변환하는 식 (4)를 이용하여 분할된 Mel값을 주파수로 역변환 후 식 (5)를 통해 48개의 Mel-scale Filter($H_i(k)$)를 갖는 Mel Filter Bank를 계산한다. 여기서 f_i 는 Mel에서 역변환된 주파수 값을 의미한다(Hossain *et al.*, 2010).

$$M = 2595 \log_{10}(1 + \frac{f}{700}) \quad (3)$$

$$f = 700(10^{\frac{M}{2595}} - 1) \quad (4)$$

$$H_i(k) = \begin{cases} 0 & k < f_{i-1} \\ \frac{k - f_{i-1}}{f_i - f_{i-1}} & f_{i-1} \leq k < f_i \\ 1 & k = f_i \\ \frac{f_{i+1} - k}{f_{i+1} - f_i} & f_i < k \leq f_{i+1} \\ 0 & k > f_{i+1} \end{cases} \quad i = 1, 2, \dots, 48 \quad (5)$$

Mel Filter Bank를 통과한 출력값에 로그 변환을 한 이후 DCT를 통하여 MFCC(C_n)의 값을 구한다. 식 (6)에서 $x[k]$ 는 k번째 Mel Scale Filter의 출력값이며, n은 추출할 MFCC의 개수로 총 20개의 MFCC를 추출한다(Davis and Mermelstein, 1980).

$$C_n = \sum_{k=1}^{48} x[k] \cos(n(k - \frac{1}{2}) \frac{\pi}{48}), n = 1, 2, \dots, 20 \quad (6)$$

식 (1) ~ (6)의 과정을 통해 추출된 MFCC는 <Figure 4>와 같다. 시간에 따른 진폭의 세기를 나타내는 1차원의 심장 소리 데이터와 특징 추출을 통해 변환된 시간과 주파수의 2차원 데이터인 MFCC를 시각화하였다.

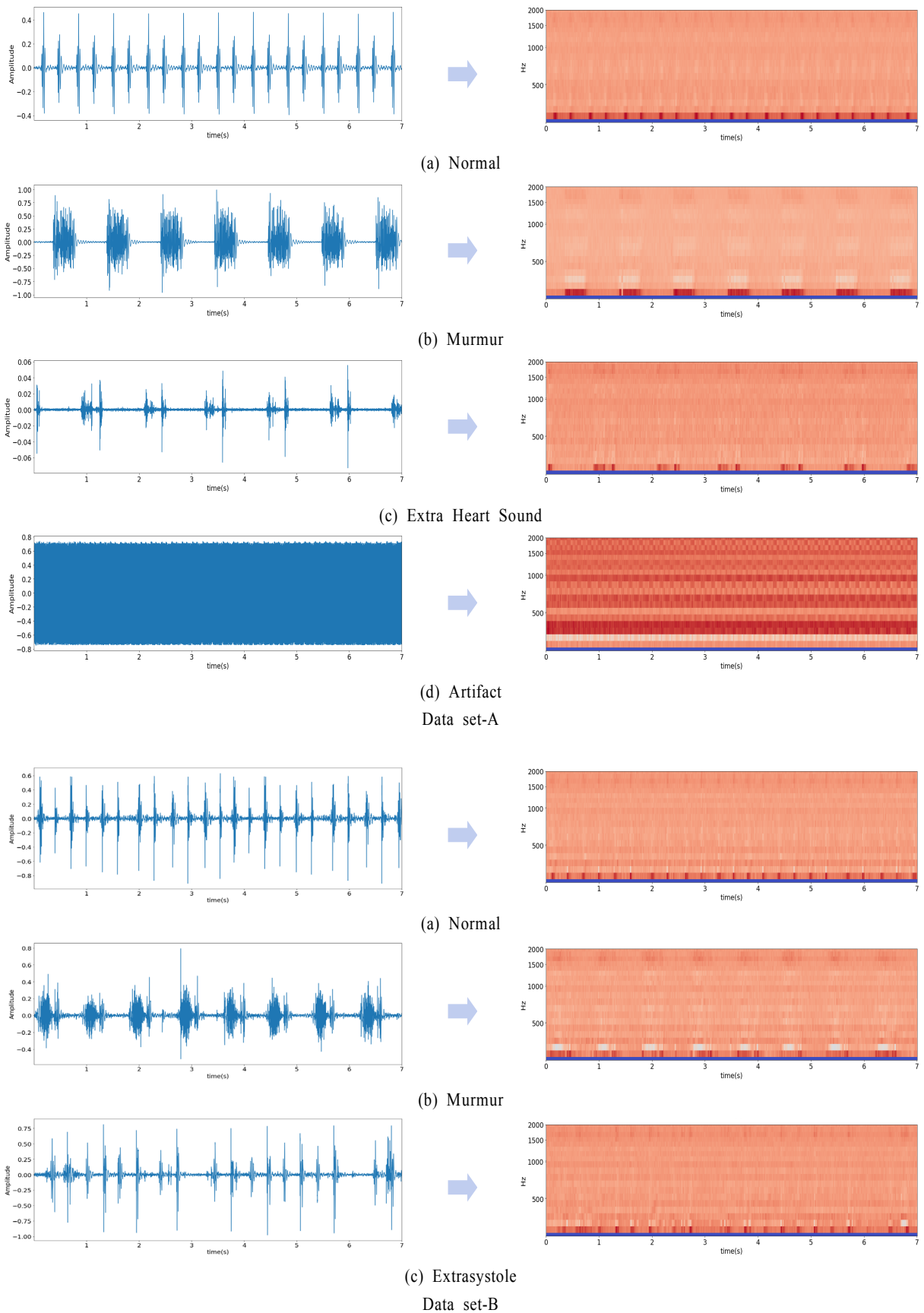


Figure 4. Heart Sound Feature Extraction : Examples of MFCC

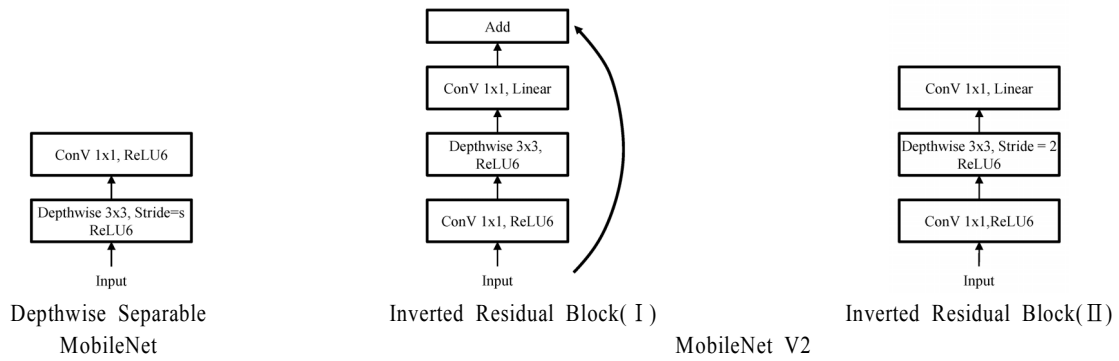


Figure 5. Comparisons on different structures of MobileNet

3.4 분류 모델

헬스케어 분야의 소형 디바이스에 적용 가능한 경량화 모델을 만들기 위해 모바일 넷 V2에서 사용한 Inverted Residual Block을 적용하였다. Inverted Residual Block은 모바일 넷(Howard *et al.*, 2017)에서 사용한 Depthwise Separable Convolution에서 비선형에 의한 정보 손실을 막기 위해 Linear bottlenecks를 적용한 구조이다. Depthwise Separable Convolution은 3×3 Filter를 채널에 독립적으로 적용한 Depthwise Convolution과 1×1 Filter를 이용하여 채널 수를 조절하는 Pointwise Convolution이 결합한 형태이다. 그리고 Inverted Residual Block은 일반적인 Residual Block과 반대로, 비선형성에 의해 손실되는 정보를 상쇄시키기 위해 Expansion Layer를 통해 채널을 확장 시키고 Depthwise Separable Convolution을 적용하며, 채널을 축소할 이후에는 비선형 함수를 사용하지 않는 구조이다. 모바일 넷과 모바일 넷 V2에서 사용한 모델 구조는 <Figure 5>와 같다.

립적으로 적용한 Depthwise Convolution과 1×1 Filter를 이용하여 채널 수를 조절하는 Pointwise Convolution이 결합한 형태이다. 그리고 Inverted Residual Block은 일반적인 Residual Block과 반대로, 비선형성에 의해 손실되는 정보를 상쇄시키기 위해 Expansion Layer를 통해 채널을 확장 시키고 Depthwise Separable Convolution을 적용하며, 채널을 축소할 이후에는 비선형 함수를 사용하지 않는 구조이다. 모바일 넷과 모바일 넷 V2에서 사용한 모델 구조는 <Figure 5>와 같다.

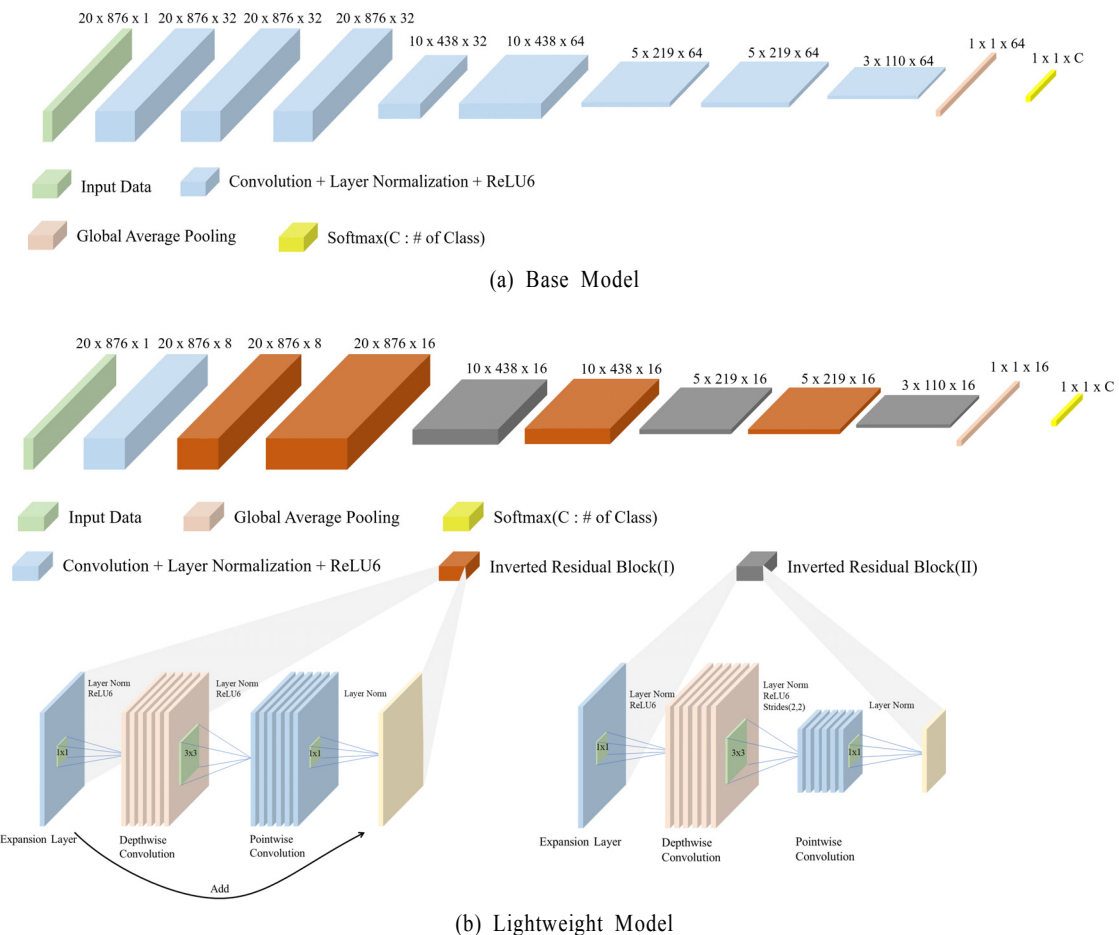


Figure 6. The architecture of the proposed CNN method

Inverted Residual Block은 2가지 종류로 구분되며, 차이는 Depthwise Convolution의 Stride와 Skip Connection의 유무이다. <Figure 5>의 Inverted Residual Block(I)는 Depthwise Convolution의 Stride를 1로 학습하여 데이터의 크기를 그대로 유지하여 입력값과 출력값을 더해주는 Skip Connection의 구조를 가진다. Inverted Residual Block(II)는 Depthwise Convolution의 Stride를 2로 학습하여 데이터의 크기를 반으로 감소시켜 모델의 크기와 연산량을 줄이는 역할을 한다.

따라서, 본 연구에서는 Inverted Residual Block을 사용한 경량화 모델(Lightweight Model)을 제안하였으며, 기본 모델(Base Model)과 비교하였다. 경량화 모델에 사용한 Inverted Residual Block은 각 Feature에 대한 평균과 분산을 통해 정규화를 하여 Batch Size에 독립적인 Layer Normalization(Ba *et al.*, 2016)을 이용하여 모델을 구성하였다. 연구에서 제안하는 기본 모델 및 경량화 모델의 구조는 <Figure 6>과 같다.

4. 실험 및 실험 결과

4.1 실험

4.1.1 실험 환경

본 연구는 윈도우 10 환경에 파이썬 3.6과 텐서플로우 2.3.1, 케라스 2.4.3을 사용하였다. 그리고 CPU는 Intel의 i7-7700, GPU는 nvidia사 GTX 1080 Ti, RAM은 64GB를 사용하였다.

4.1.2 하이퍼 파라미터 설정

Grid Search를 통해 선정한 하이퍼 파라미터는 다음과 같다. Batch Size는 8로 설정하고, Adam Optimizer를 사용하였다. 손실 함수는 Cross Entropy를 사용하였으며, 학습률(Learning Rate)은 Cosine Annealing을 사용하여 특정 주기마다 학습률을 초기화하며 학습을 하였으며 최초 학습률은 0.001로 설정하였다(Loshchilov and Hutter, 2016).

4.1.3 평가지표 선정

본 연구에서 사용하는 평가지표는 Precision, Sensitivity, Specificity, Youden's Index, Discriminant Power와 Accuracy이다. Accuracy를 제외한 평가지표는 경연대회에서 지정한 지표이며, Accuracy는 본 연구에서 추가한 평가지표이다. 경연대회에서 지정한 평가지표를 계산하기 위하여 경연대회 공식 사이트에서 엑셀 파일을 제공하고 있으며, 파일에는 Data set A와 B에 해당하는 시트가 존재한다(Bentley *et al.*, 2011). 각 시트에 평가 데이터를 분류한 결과를 One-hot-encoding을 하여 입력하면 클래스별로 예측한 개수와 그 중 올바르게 예측한 개수를 계산하여 미리 작성된 함수에 의해 자동으로 평가지표를 산출한다. 이에 식 (7) ~ (19)는 엑셀 파일에 주어진 평가지표에 대

한 수식이다.

1. **Precision** (Pr)은 모델이 정답으로 분류한 데이터 중 실제 정답인 것의 비율이다. 다중 분류 문제에서는 모델이 특정 클래스로 분류한 데이터 중 실제로 해당 클래스인 것의 비율을 의미한다.

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

본 연구에서는 Data set A에서 Normal, Murmur, Extra Heart Sound, Artifact의 Precision과 4개 클래스의 Precision의 합인 Total Precision을 계산한다. Data set B에서도 클래스들의 Precision과 Precision들의 합인 Total Precision을 산출한다.

$$Total\ Precision_{Data\ set\ A} = Pr_{Normal} + Pr_{Murmur} + Pr_{ExtraHS} + Pr_{Artifact} \quad (8)$$

$$Total\ Precision_{Data\ set\ B} = Pr_{Normal} + Pr_{Murmur} + Pr_{Extrasystole} \quad (9)$$

PASCAL 경연대회에서 지정된 Data set A의 평가지표는 Artifact 클래스의 Sensitivity와 Specificity이며, Data set B에서는 비정상 심장 소리에 관련된 클래스의 Sensitivity와 Specificity를 사용한다.

2. **Sensitivity** (*Se*)는 실제 정답을 모델이 정답으로 올바르게 예측하는 비율이다. 다중 분류 문제에서 특정 클래스의 Sensitivity는 해당 클래스의 데이터를 모델이 올바르게 예측한 비율을 의미한다.

$$Sensitivity = \frac{TP}{TP+FN} \quad (10)$$

3. **Specificity** (*Sp*)는 실제 정답이 아닐 때 정답이 아니라고 예측하는 비율을 의미한다. 이를 다중 분류 문제에 적용하면 특정 클래스의 Specificity는 해당 클래스를 제외한 클래스의 데이터를 모델이 올바르게 예측한 비율이다.

$$Specificity = \frac{TN}{FP+TN} \quad (11)$$

4. **Youden's Index**는 Sensitivity와 Specificity가 최대가 되는 점을 구하는 지표이다. 이는 분류 문제에서 최적의 Cut-off 지점을 찾는 방법으로 주로 사용된다. Youden's Index는 Data set A와 B에서 사용하며, Artifact와 비정상 심장 소리에 관련된 클래스의 Sensitivity와 Specificity를 사용한다.

$$Youden's\ Index = Se - (1 - Sp) \quad (12)$$

5. **Discriminant Power**(Dp)는 분류 모델이 클래스의 궁·부정 사례를 얼마나 잘 분류하는지 측정하는 지표이다. Discriminant Power는 Data set B에서 Normal 클래스와 Murmur, Extrasystole 클래스의 분류 정도를 측정하기 위해 사용한다.

$$Dp = \frac{\sqrt{3}}{\pi} (\log \frac{Se}{1-Se} + \log \frac{Sp}{1-Sp}) \quad (13)$$

6. **Accuracy**(정확도)는 전체 데이터 중 올바르게 분류된 데이터의 비율을 의미한다. Accuracy는 경연대회에서 사용한 Sensitivity와 Specificity를 기준으로 계산하였다.

$$Accuracy_{Data\ set\ A} = \frac{16 \times Se_{Artifact} + 36 \times Sp_{Artifact}}{52} \quad (14)$$

$$Accuracy_{Data\ set\ B} = \frac{59 \times Se_{Heart\ problem} + 136 \times Sp_{Heart\ problem}}{195} \quad (15)$$

식 (14)의 $Se_{Artifact}$ 와 $Sp_{Artifact}$ 는 식 (16), (17)과 같이 계산되므로 $16 \times Se_{Artifact}$ 는 Artifact의 TP를, $36 \times Sp_{Artifact}$ 는 Artifact를 제외한 클래스들(Normal, Murmur, Extra HS)의 TP의 합을 의미한다. Data set A의 4개 클래스의 TP의 합을 평가 데이터의 수로 나누어 Accuracy를 산출한다.

$$Se_{Artifact} = \frac{TP_{Artifact}}{16}, \quad Data\ set\ A \quad (16)$$

$$Sp_{Artifact} = \frac{TP_{Normal} + TP_{ExtraHS} + TP_{Murmur}}{36}, \quad Data\ set\ A \quad (17)$$

식 (15)의 $Se_{Heart\ problem}$ 와 $Sp_{Heart\ problem}$ 는 식 (18), (19)와 같이 계산되어 $59 \times Se_{Heart\ problem}$ 은 Normal을 제외한 클래스들(Murmur, Extrasystole)의 TP의 합을, $136 \times Sp_{Heart\ problem}$ 은 Normal 클래스의 TP를 의미한다. Data set B의 3개 클래스의 TP의 합을 평가 데이터의 수로 나누어 Accuracy를 산출한다.

$$Se_{Heart\ problem} = \frac{TP_{Murmur} + TP_{Extrasystole}}{59}, \quad Data\ set\ B \quad (18)$$

$$Sp_{Heart\ problem} = \frac{TP_{Normal}}{136}, \quad Data\ set\ B \quad (19)$$

4.2 실험 결과

4.2.1 기본 모델과 경량화 모델 성능 평가

5-Fold Cross Validation을 통해 모델을 검증하였으며, 실험 결과는 <Table 2>에 Total Precision과 Accuracy로 요약하였다. Precision은 모델이 특정 클래스로 분류한 데이터 중 실제로 해당 클래스인 것의 비율을 의미하기 때문에, 특정 클래스로 예측된 데이터의 수가 적지만, 예측된 데이터가 실제 분류와 일치하는 경우, 해당 클래스의 Precision이 높아지는 경우가 발생한다. 이에, 경연대회 평가지표인 Total Precision에 더하여 Accuracy를 통해 결과를 해석하였다. Data set A의 경량화 모델의 4-Fold는 Total Precision이 2.68로 3, 5-Fold에 비해 낮지만, Accuracy는 0.71로 3-Fold와 5-Fold에 비해 2% 높다. 그리고 Data set B의 기본 모델의 3-Fold ~ 5-Fold의 결과를 보면 Total Precision의 결과는 차이가 있지만, 정확도는 동일한 것을 확인할 수 있다.

Table 2. Classification Performance in Each Fold

		Base Model		Lightweight Model	
		Total Precision	Accuracy	Total Precision	Accuracy
Data set A	1-Fold	3.04	0.77	2.92	0.75
	2-Fold	2.75	0.71	2.73	0.71
	3-Fold	2.69	0.69	2.64	0.67
	4-Fold	2.97	0.73	2.68	0.71
	5-Fold	2.78	0.71	2.7	0.69
	Average	<u>2.84</u>	<u>0.722</u>	<u>2.734</u>	<u>0.706</u>
Data set B	1-Fold	2.44	0.83	2.21	0.81
	2-Fold	1.98	0.83	2.09	0.81
	3-Fold	2.19	0.82	1.93	0.81
	4-Fold	2.23	0.82	1.92	0.78
	5-Fold	1.93	0.82	1.87	0.79
	Average	<u>2.154</u>	<u>0.824</u>	<u>2.004</u>	<u>0.8</u>

Table 3. Results on Data set A

Data set A	Zhang, 2017a	Zhang, 2017b	Chakir, 2018	Bourouhou, 2020	Proposed Method	
					Base Model	Lightweight
Precision of Normal	0.67	0.61	0.59	0.7	0.86	0.85
Precision of Murmur	1	0.91	0.82	0.63	0.9	0.75
Precision of ExtraHS	0.43	0.5	0.83	1	0.57	0.57
Precision of Artifact	0.8	0.94	0.72	0.73	0.71	0.75
Artifact Sensitivity	1	1	0.81	1	0.94	0.94
Artifact Specificity	0.64	0.67	0.67	0.56	0.69	0.67
Youden's Index	0.64	0.67	0.48	0.56	0.63	0.6
Total Precision	2.9	2.96	2.96	3.06	3.04	2.92
Accuracy	0.75	0.77	0.71	0.7	0.77	0.75

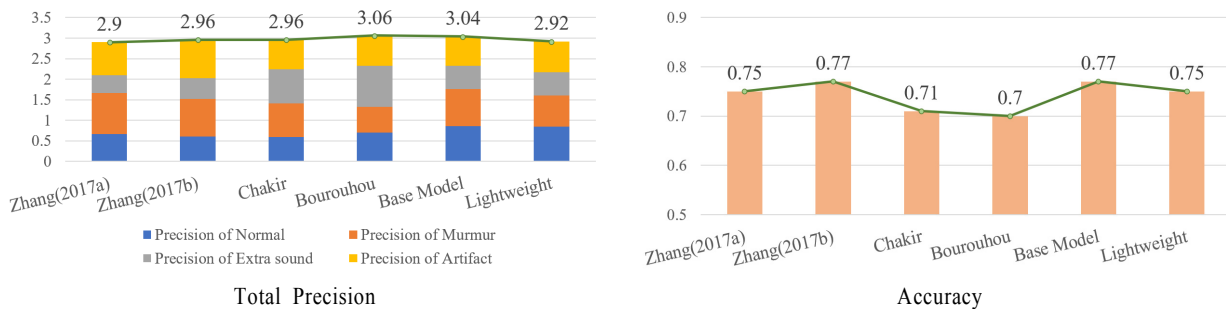


Figure 7. Summary of The Classification Performance on Data set A

Table 4. Results on Data set B

Data set B	Zhang, 2017a	Zhang, 2017b	Chakir, 2018	Bourouhou, 2020	Proposed Method	
					Base Model	Lightweight
Precision of Normal	0.83	0.81	0.82	0.76	0.81	0.81
Precision of Murmur	0.7	0.67	0.59	0.61	0.96	0.89
Precision of Extrasystole	0.15	0.14	0.18	1	0.67	0.5
Sensitivity of heart problem	0.49	0.51	0.49	0.36	0.47	0.47
Specificity of heart problem	0.84	0.8	0.66	0.9	0.99	0.96
Youden's Index	0.33	0.31	0.15	0.26	0.46	0.44
Discriminant Power	0.39	0.34	0.15	0.4	0.98	0.76
Total Precision	1.68	1.62	1.58	2.37	2.44	2.21
Accuracy	0.73	0.71	0.61	0.74	0.83	0.81

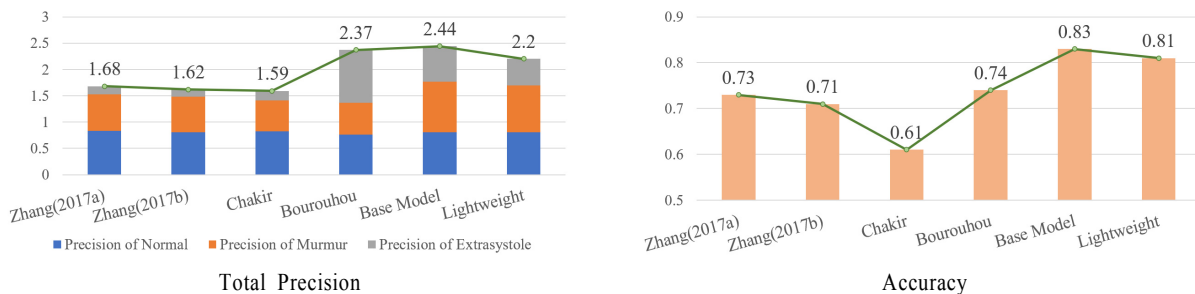


Figure 8. Summary of The Classification Performance on Data set B

4.2.2 선행연구 비교

<Table 3>, <Table 4>에서 제시한 성능은 5개 Fold 중 제일 좋은 성능이 나온 결과를 제시하였다.

Data set A에서 기본 모델과 경량화 모델 결과는 <Table 3>과 같다. Normal 클래스의 Precision이 선행연구에 비해 우수한 성능을 보였으며, Total Precision은 선행연구들과 유사한 성능을 보이는 것을 확인하였다. 그러나 Precision의 평가지표의 한계에 따라, Accuracy를 비교하여 확인하면, Bourouhou, 2020은 Total Precision을 기준으로 성능이 가장 우수하였지만, Accuracy는 다른 선행연구 및 제안한 모델에 비해 낮은 것을 확인하였다. 반면, 제안한 경량화 모델은 선행연구에 비해 Total Precision은 2.92로 낮은 성능을 보였지만, Accuracy는 0.75로 높은 성능을 보였다. 그리고 기본 모델의 Artifact 클래스에 대한 Specificity는 0.69로 선행연구에 비해 높은 것을 확인하였으며, 경량화 모델은 0.67로 선행연구들과 유사한 성능을 유지하는 것을 확인하였다. Artifact는 심장 소리 이외의 언어, 음악 및 소음을 포함한 클래스로 Artifact에 대한 Specificity는 모델이 Artifact를 제외한 심장의 이상 유무와 연관된 소리의 분류 정확도를 측정하는 지표이다. 이에, 제안한 모델이 심장 소리 이상 유무를 확인할 수 있는 클래스들에 대하여 선행연구에 비해 더 좋은 설명력을 가짐을 검증하였다. Data set A의 Total Precision과 Accuracy를 요약한 결과는 <Figure 7>과 같다.

Data set B에서 기본 모델과 경량화 모델의 결과는 <Table 4>와 같다. 기본 모델과 경량화 모델 모두 Murmur의 Precision은 선행연구에 비해 우수한 성능을 보였으며, Normal의 Precision은 선행연구와 비슷한 성능을 보이는 것을 확인하였다. 그리고 제안한 모델이 heart problem의 Sensitivity는 선행연구와 비슷한 성능을, Specificity와 Youden's Index, Discriminant Power에서 높은 성능을 보여주었다. 이 4가지 지표는 정상 심장 소리와 이상 심장 소리의 분류 정도를 나타내는 지표로, 제안하는 모델이 심장 소리의 이상 유무를 잘 분류하고 있음을 확인할 수 있었다. 그리고 Total Precision 기준으로 기본 모델은 2.44로 가장 높은 성능을 보였으며, 경량화 모델은 소폭 하락하여 2.21의 성능을 보였으나 Accuracy 기준으로 기본 모델은 0.83, 경량화 모델은 0.81로 선행연구에 비해 높은 성능을 보여 모델의 우수성을 검증하였다. Data set B의 Total Precision과 Accuracy를 요약한 결과는 <Figure 8>과 같다.

선행연구들은 머신러닝 기법(SVM, KNN, Tree 기반 등)을 제안하였으며, Zhang, 2017b는 스펙트로그램을 추출한 후 텐서 분해를 통해 스케일링하여 CNN 분류 모델의 입력 데이터로 사용한 딥러닝 기법(WSCNN)을 제안한 방법론이다. 이에 <Table 5>는 Zhang, 2017b의 연구와 제안하는 기본 모델과 Inverted Residuals 구조를 적용한 경량화 모델의 파라미터 및 성능을 비교한 표이다.

기본 모델은 선행연구와 파라미터에서 큰 차이가 없으나, Data set A에서 Total Precision 기준 약 2.7%의 성능이 향상되었으며, Accuracy는 동일한 성능을 보였다. Data set B에서 Total Precision은 약 50.6%, Accuracy 기준 약 16.9%의 성능이 향상되었다.

경량화된 모델은 선행연구에 비해 약 82.5% 적은 파라미터를 학습하며, Data set A에서 Total Precision 기준 약 1.4%의 성능이 감소하였으며, Accuracy는 약 2.6%로 소폭 하락하였다. Data set B에서 Total Precision은 약 36.4%, Accuracy 기준 약 14.1%의 성능이 향상되었다.

제안한 모델과 Zhang, 2017b 연구의 큰 차별점은 모델을 분류하기 위해 사용한 특징과 사용한 모델의 구조이다. Zhang은 스펙트로그램을 특징으로 사용하였으나, 심장 소리는 주로 낮은 주파수 대역대에 분포하므로 본 연구에서 사용한 MFCC가 스펙트로그램에 비해 심장 소리의 특징 추출에 더 적합하였다. 그리고 Total Precision과 Accuracy를 통해 제안한 Inverted Residual Block 모델이 더 적은 파라미터로 기존 모델의 성능을 유지할 수 있는 것을 확인하여 모델 구조의 효율성을 입증하였다.

4.2.3 경량화 비교 연구

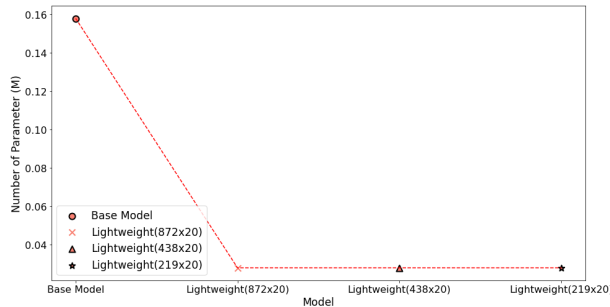
헬스케어 디바이스에 탑재하기 위한 경량화 모델의 성능을 정량적으로 평가하기 위하여 모델의 입력 데이터의 크기를 조절하여 비교 실험을 하였다. 입력 데이터의 크기는 MFCC 추출 시 Framing 과정에서 Frame의 길이를 기존에 비해 2배, 4배로 조절하여 입력 데이터의 크기를 1/2, 1/4로 감소시켰다. 이를 통해 학습 파라미터의 수, 모델 파일 크기, 연산량(MAdds), 추론 시간, 모델 성능(Total Precision, Accuracy)을 비교하여 경량화 성능을 평가하였다.

Table 5. Comparison of Models with respect to Size and Performance

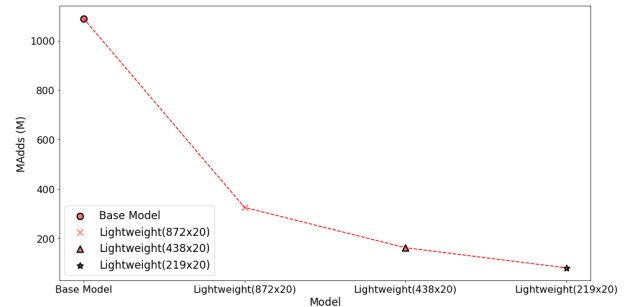
		Zhang, 2017b	Base Model		Lightweight Model	
			Performance	Improvement Ratio	Performance	Improvement Ratio
The number of Parameters		160K	158K	- 1.25%	28K	- 82.5%
Data set A	Total Precision	2.96	3.04	+ 2.7%	2.92	- 1.4%
	Accuracy	0.77	0.77	.	0.75	- 2.6%
Data set B	Total Precision	1.62	2.44	+ 50.6%	2.21	+ 36.4%
	Accuracy	0.71	0.83	+ 16.9%	0.81	+ 14.1%

Table 6. Comparison between Base Model and Lightweight Models

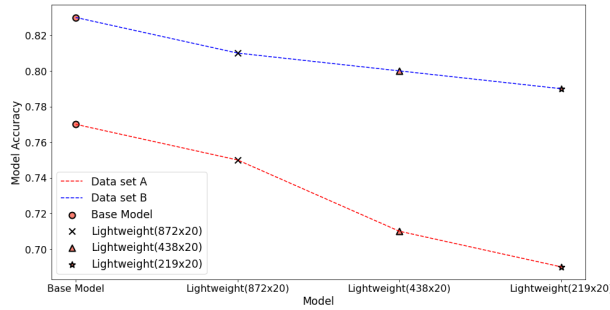
Model		Base Model	Lightweight Models		
Input Size		872×20	872×20	438×20	219×20
The number of Parameters (M)		0.158	0.028		
Size of Model (Kb)		2,009	731		
MAdds (M)		1,090	326	163	81.6
Data set A	Inference Time (ms)	67	57	53	51
	Total Precision	3.04	2.92	2.85	2.71
	Accuracy	0.77	0.75	0.71	0.69
Data set B	Inference Time (ms)	70	59	53	52
	Total Precision	2.44	2.21	2.16	2.00
	Accuracy	0.83	0.81	0.8	0.79



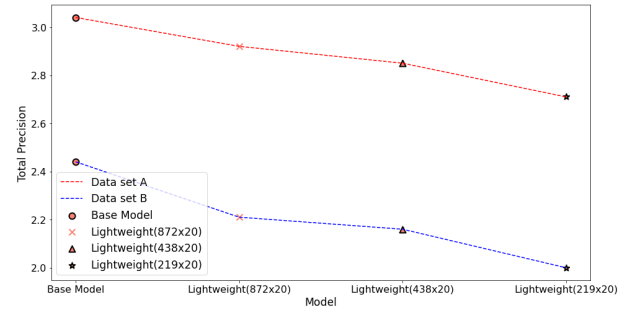
(a) Comparison of Parameters



(b) Comparison of MAdds



(c) Comparison of Accuracy



(d) Comparison of Total Precision

Figure 9. The graphical results of the Base Model and Lightweight Models

학습 파라미터의 수는 기본 모델은 약 158,000개이며, 경량화 모델은 약 28,000개로 82% 가량 감소하였다. 모델 파일의 크기는 학습 파라미터의 수에 비례하여, 기본 모델이 2,009Kb이며 경량화 모델이 731Kb로 약 73% 감소하였다. 경량화 연구에서 연산량을 평가하기 위해 MAdds (Multiply-Adds)를 사용한다. MAdds는 모델의 연산량을 나타내는 대표적인 값으로 Convolution layer에서의 연산량은 식 (20)을 통해 계산된다. $D_K \times D_K$ 는 필터의 크기이며, N 은 필터의 개수, M 은 입력 데이터의 채널 수, $D_F \times D_F$ 는 Feature map의 크기를 의미한다 (Howard *et al.*, 2017).

$$D_K \times D_K \times N \times M \times D_F \times D_F \quad (20)$$

기본 모델과 같은 크기의 입력 데이터를 사용하는 경량화 모델의 연산량은 기본 모델 대비 약 70%가량 감소하였으며, 경량화 모델 내에서는 입력 데이터의 크기에 비례하여 감소하였다. 그리고 입력 데이터의 크기와 파라미터의 개수에 따라 모델의 성능을 확인할 수 있으며 특히, 데이터의 개수가 적은 Data set A에서 더 큰 폭으로 감소하였으며, Data set B에서는 입력 데이터의 크기가 감소하여도 성능을 유지하는 것을 확인할 수 있었다. 모바일 넷 V2의 실험 결과에 의하면 분류 문제에서 3.4M 파라미터와 300M의 MAdds를 사용하여 모델을 학습하였고 이를 모바일 기기에 탑재하여 실행 시 약 75ms가 소요되었다고 한다(Sandler *et al.*, 2018). 이를 통해 본 연구에서 제안하는 모델은 파라미터와 연산량을 기준으로 소형 헬스케

어 기기에 탑재가 가능하다고 할 수 있다.

추론 시간은 Data set A와 B의 평가 데이터 개수의 차이가 있으므로 평가 데이터 중 30개를 임의로 추출하여 전처리, 특징 추출, 모델 로드 및 모델 추론까지의 평균 시간을 측정하였다. 연산량이 적을수록 추론 시간이 빨라짐을 알 수 있다. <Figure 9>를 통해 Model에 따른 학습 파라미터 수와 연산량, Accuracy와 Total Precision을 나타내었다.

5. 결론

본 연구는 PASCAL 데이터를 사용하여 심장 소리를 분할하지 않고 이상 여부를 분류할 수 있으며, 헬스케어 디바이스에 탑재할 수 있는 모바일 넷 V2의 Inverted Residual Block 기반 경량화 모델을 제안하였다. 연구 방법으로는 Band Pass Filter를 사용하여 잡음을 제거하고 Repeat 기법을 통해 모든 데이터를 동일한 길이로 맞춰주며, 심장 소리에 적합한 특징을 추출하기 위하여 MFCC를 사용하였다. 연구에 대한 공정성을 가지기 위하여 선행연구와 동일한 라벨링 되어 있지 않은 평가 데이터를 사용하여 본 연구에서 제안하는 경량화 모델의 우수성을 입증하였다. 또한 경량화 비교 실험을 통해 학습 파라미터 개수와 연산량, 모델 파일 크기, 추론 시간 및 모델 정확도를 이용하여 경량화의 우수성을 확인하였다. 본 연구에서 제안한 Inverted Residual Block 기반의 경량화된 심장 소리 분류 모델은 코로나19와 같은 팬데믹이 가져온 비대면 상황에서 지속적인 질량 관리의 모니터링을 비롯하여 소외계층의 진료 접근성에 도움을 줄 것으로 기대한다.

본 연구의 제한점은 모바일 및 임베디드 환경에 탑재 가능한 경량화 모델을 제시하였으나, 실제 제한된 환경을 구성하는 데 한계가 있었다. 이에 향후 연구에서는 추가적인 임상 데이터 확보와 실제 모바일 환경을 구성하여 실생활에 적용 가능한 경량화된 모델을 제안할 수 있을 것이다.

참고문헌

- Ba, J. L., Kiros, J. R., and Hinton, G. E. (2016), Layer normalization, arXiv preprint arXiv:1607.06450.
- Balili, C. C., Sobrepena, M. C. C., and Naval, P. C. (2015), Classification of Heart Sounds Using Discrete and Continuous Wavelet Transform and Random Forests, *2015 3rd IAPR Asian conference on pattern recognition (ACPR)*, IEEE, 655-659.
- Bentley, P., Nordehn, G., Miguel, C., Mannor, S., and GetZ, R. (2011), <http://www.peterjbentley.com/heartchallenge/>.
- Bourouhou, A., Jilbab, A., Nacir, C., and Hammouch, A. (2020), Heart Sound Signals Segmentation and Multiclass Classification, *International Association of Online Engineering*.
- Chakir, F., Jilbab, A., Nacir, C., and Hammouch, A. (2018), Phonocardiogram signals processing approach for PASCAL classifying heart sounds challenge, *Signal, Image and Video Processing*, 12(6), 1149-1155.
- Davis, S. and Mermelstein, P. (1980), Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(4), 357-366.
- Deng, Y., and Bentley, P. J. (2012), A robust heart sound segmentation and classification algorithm using wavelet decomposition and spectrogram, *Workshop Classifying Heart Sounds*, La Palma, Canary Islands, 1-6.
- Deng, S. W., and Han, J. Q. (2016), Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps, *Future Generation Computer Systems*, 60, 13-21.
- Deng, M., Meng, T., Cao, J., Wang, S., Zhang, J. and Fan, H. (2020), Heart sound classification based on improved MFCC features and convolutional recurrent neural networks, *Neural Networks*, 130, 22-32.
- Gomes, E. F., and Pereira, E. (2012), Classifying heart sounds using peak location for segmentation and feature construction, *Workshop Classifying Heart Sounds*, 480-492.
- Ha, T.-Y. and Lee, H.-J. (2020), Implementation of Application for Smart Healthcare Exercise Management Based on Artificial Intelligence, *The Institute of Electronics and Information Engineers*, 57(6), 44-51.
- Hossain, M. A., Memon, S., and Gregory, M. A. (2010), A novel approach for MFCC feature extraction, *2010 4th International Conference on Signal Processing and Communication Systems*, 1-5.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017), Mobilenets: Efficient convolutional neural networks for mobile vision applications, arXiv preprint arXiv:1704.04861
- Kang, S.-H, Trung, P. T., Lee, H.-K, Cho, S.-W., and Lee, K.-S. (2020), Hazardous Sound Detection System Using Modified MFCC, *Journal of Korean Institute of Intelligent Systems*, 30(1), 47-53.
- Kwon, S. -S. (2021), Utilization Assignment of Medical Information Healthcare in The 4th Industrial Revolution Era, *Journal of Culture Industry*, 21(2), 119-124.
- Li, T., Yin, Y., Ma, K., Zhang, S. and Liu, M. (2021), Lightweight End-to-End Neural Network Model for Automatic Heart Sound Classification, *Information, Multidisciplinary Digital Publishing Institute*, 12(2), 54.
- Liu, C., Springer, D., Li, Q., Moody, B., Juan, R. A., Chorro, F. J., Castells, F., Roig, J. M., Silva, I., and Johnson, A. E. W. (2016), An open access database for the evaluation of heart sound algorithms, *Physiological Measurement*, 37(12), 2181.
- Loshchilov, I. and Hutter, F. (2016), Sgdr: Stochastic gradient descent with warm restarts, arXiv preprint arXiv:1608.03983
- Oppenheim, A. V (1999), *Discrete-time signal processing*, Pearson Education India
- Raza, A., Mehmood, A., Ullah, S., Ahmad, M., Choi, G.-S., and On, B.-W. (2019), Heartbeat sound signal classification using deep learning, *Sensors, Multidisciplinary Digital Publishing Institute*, 19(21), 4819.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L. C. (2018), Mobilenetv2: Inverted residuals and linear bottlenecks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4510-4520.
- Shuvo, S. B., Ali, S. N., Swapnil, S. I., Al-Rakhami, M. S., and Gumaei, A. (2021), CardioXNet: A novel lightweight deep learning framework for cardiovascular disease classification using heart

- sound recordings, *IEEE*, 9, 36955-36967.
- Vrbancic, G. and Podgorelec, V. (2019), Automatic detection of heartbeats in heart sound signals using deep convolutional neural networks, *Elektronika ir Elektrotechnika*, 25(3), 71-76.
- Xiao, B., Xu, Y., Bi, X., Zhang, J., and Ma, X. (2020), Heart sounds classification using a novel 1-D convolutional neural network with extremely low parameter consumption, *Neurocomputing*, Elsevier 392, 153-159.
- Zhang, W., Han, J., and Deng, S. (2017), Heart sound classification based on scaled spectrogram and partial least squares regression, *Biomedical Signal Processing and Control*, 32, 20-28.
- Zhang, W., Han, J., and Deng, S. (2017a), Heart sound classification based on scaled spectrogram and tensor decomposition, *Expert Systems with Applications*, 84, 220-231.
- Zhang, W. and Han, J. (2017b), Towards heart sound classification without segmentation using convolutional neural network, *2017 Computing in Cardiology (CinC)*, IEEE, 1-4

저자소개

박두서: 육군사관학교 수학과에서 2013년 학사학위를 취득하고, 고려대학교 산업경영공학과 석사과정에 재학 중이다. 연구 분야는 Healthcare, 딥러닝이다.

이민영: 고려대학교 산업경영공학부에서 2014년 학사학위를 취득하고 고려대학교 산업경영공학과에서 2021년 박사학위를 취득하였다. 현재 카카오브레인 연구원으로 재직하고 있다. 연구 분야는 Combinatorial Optimization, AI이다.

김기현: 고려대학교 경영정보공학과에서 2011년 학사학위를 취득하고 고려대학교 산업경영공학과에서 2015년 석박통합과정을 수료하였다. 연구 분야는 Text Mining, NLP, Chatbot(Intent & Entity)이다.

이홍철: 고려대학교 산업공학부에서 1983년 학사, Texas Arlington 대학교 산업공학과에서 1988년 석사학위를 취득하고 Texas A&M 대학교에서 산업공학 박사학위를 취득하였다. 1996년부터 고려대학교 산업경영공학부 교수로 재직하고 있다. 연구 분야는 AI, 생산공학시스템, 시뮬레이션이다.