

# 다양한 합성곱 신경망 방식을 이용한 폐음 분류 방식의 성능 비교

## Performance comparison of lung sound classification using various convolutional neural networks

김지연,<sup>1</sup> 김형국<sup>†</sup>

(Gee Yeun Kim<sup>1</sup> and Hyoung-Gook Kim<sup>1†</sup>)

<sup>1</sup>광운대학교 전자융합공학과

(Received April 11, 2019; accepted June 19, 2019)

**초 록:** 폐질환 진단에서 청진은 다른 진단 방식에 비해 단순하고, 폐음을 이용하여 폐질환 환자식별뿐 아니라 폐음과 관련된 질병을 예측할 수 있다. 따라서 본 논문에서는 다양한 합성곱 신경망 방식을 기반으로 폐음을 이용하여 폐질환 환자를 식별하고, 소리특성에 따른 폐음을 분류하여 각 신경망 방식의 분류 성능을 비교한다. 먼저 폐질환 소견을 갖는 흉부 영역에서 단채널 폐음 녹음기기를 이용하여 폐음 데이터를 수집하고, 수집된 시간축 신호를 스펙트럼 형태의 특징값으로 추출하여 각 분류 신경망 방식에 적용한다. 폐 사운드 분류 방식으로는 일반적인 합성곱 신경망, 병렬 구조, 잔류학습이 적용된 구조의 합성곱 신경망을 사용하고 실험을 통해 각 신경망 모델의 폐음 분류 성능을 비교한다.

**핵심용어:** 폐음 분류, 병렬 합성곱 신경망, 잔류 학습, 성능 비교

**ABSTRACT:** In the diagnosis of pulmonary diseases, auscultation technique is simpler than the other methods, and lung sounds can be used for predicting the types of pulmonary diseases as well as identifying patients with pulmonary diseases. Therefore, in this paper, we identify patients with pulmonary diseases and classify lung sounds according to their sound characteristics using various convolutional neural networks, and compare the classification performance of each neural network method. First, lung sounds over affected areas of the chest with pulmonary diseases are collected by using a single-channel lung sound recording device, and spectral features are extracted from the collected sounds in time domain and applied to each neural network. As classification methods, we use general, parallel, and residual convolutional neural network, and compare lung sound classification performance of each neural network through experiments.

**Keywords:** Lung sound classification, Parallel convolutional neural network, Residual learning, Performance comparison

**PACS numbers:** 43.60.Bf, 43.80.Qf

### 1. 서 론

폐질환은 세계에서 10대 사망 원인 중 하나로 2015년 세계보건기구(World Health Organization, WHO)에 따르면 만성폐쇄성질환(Chronic Obstruction Pulmo-

nary Disease, COPD)으로 인해 317만명이 사망했으며, 호흡기 감염으로 인한 사망자는 319만 명이라고 보고되었다. 폐질환은 개인의 다양한 생리적 요인, 대기오염과 같은 환경의 변화로 인해 폐 염증, 손상과 같은 이상 문제로 인해 발생되며 이러한 이상 증상은 X-선 및 CT 촬영, 폐활량 측정, 청진 등과 같은 방식으로 진단할 수 있다. 그중 청진은 비외과적이고, 낮은 비용으로 진단을 가능하게 하므로 가장 중요하

<sup>†</sup>Corresponding author: Hyoung-Gook Kim (hkim@kw.ac.kr)  
Department of Electronics Convergence Engineering, Kwangwoon University, 20 Gwangun-ro, Nowon-gu, Seoul 01897, Republic of Korea  
(Tel: 82-2-940-5574, Fax: 82-2-913-5006)

면서도 기본이 되는 검사이다. 또한 폐음은 정상적인 소리, 부잡음(adventitious sound)으로 구분되어 청진을 통한 진단 방식에서 폐질환 환자식별뿐 아니라 발생한 폐음과 관련된 질병을 예측하는데 도움을 줄 수 있다. 하지만 청진은 사람의 청각에 의존하므로 의사의 충분한 수련 과정을 필요로 하고, 사람이 소리를 느낄 수 있는 가청주파수 대역은 한정되기 때문에 낮은 주파수 대역에서 발생한 폐음을 식별하는 것은 쉽지 않다. 따라서 이러한 한계를 보완하기 위해 이전 연구에서는 다양한 디지털 신호처리 방식을 이용한 폐음 분석에 대한 접근 방식을 제안하였고,<sup>[1,2]</sup> 최근에는 k-최근접이웃 알고리즘(k-Nearest Neighbor Algorithm, k-NN), 인공 신경망과 같은 인공지능 기반의 사운드 분석 및 분류 방식을 이용하여 폐음 분류 성능을 개선시켜 오고 있다.<sup>[3,4]</sup>

인공 신경망의 발전된 형태인 합성곱 신경망(Convolutional Neural Network, CNN)은 입력 이미지의 특징을 추출하는 여러 신경망의 가중치를 공유하는 방식으로 영상 처리, 오디오 신호 처리 등과 같은 다양한 분야에서 사용되고 있으며 폐음 분류 방식에도 적용되어 우수한 성능을 보이고 있다.<sup>[5]</sup> 최근에는 CNN 방식을 기반으로 분류 및 인식 성능을 향상시키기 위해 병렬 구조의 합성곱 신경망,<sup>[6]</sup> 잔류 학습을 적용한 심층 신경망<sup>[7]</sup>에 대한 연구가 활발히 진행되고 있다.

이에 본 논문에서는 한 호흡 주기에 녹음된 단채널 폐음의 스펙트럼과 시간 및 공간 정보를 기반으로 다양한 구조의 합성곱 신경망 방식인 일반적인 CNN, 병렬 CNN(Parallel CNN, P-CNN), Residual CNN(R-CNN)을 이용하여 건강한 사람과 폐질환 환자를 식별하고, 더 나아가 소리 특징에 따른 건강한 폐음, 부잡음인 수포음(crackle), 천명음(wheezing), 협착음(stridor)을 분류하여 각 신경망 방식의 성능을 비교한다.

본 논문의 구성은 다음과 같다. II장에서는 폐음 분류를 위한 다양한 합성곱 신경망 방식을 설명하고, III장에서는 폐 사운드 데이터베이스, 실험 환경 및 결과를 서술한다. 마지막으로 IV장에서는 결론과 향후 연구를 설명한다.

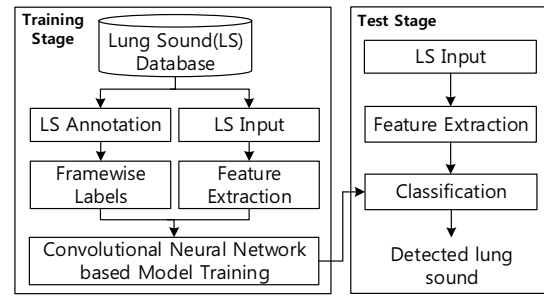


Fig. 1. Architecture of the basic framework for lung sound classification.

## II. Convolutional Neural Networks

Fig. 1은 다양한 구조의 CNN 방식을 적용하기 위한 기본적인 폐음 분류 프레임워크를 보여준다.

제안하는 프레임워크는 학습 단계와 테스트 단계로 구성된다. 학습 단계에서는 먼저 단채널 폐 사운드 녹음기기를 통해 수집된 폐 사운드의 특징을 추출하기 위해 32ms 크기의 헤밍 윈도우와 12ms 오버래핑을 적용하여 폐 사운드를 프레임 단위로 분할하고, 단시간 푸리에 변환에 적용하여 257-bin 스펙트럼을 추출한다. 추출된 2차원의 스펙트럼은 합성곱 기반의 신경망 방식에 입력되어 학습을 수행한다. 테스트 단계에서는 새로운 폐음 데이터의 특징값을 추출하고, 추출된 스펙트럼을 학습된 신경망에 적용하여 폐 사운드 분류한다. 본 논문에서는 학습된 각 신경망 모델의 분류 정확도를 통해 분류 방식의 성능을 비교하고, 폐음 분류 방식으로 일반적인 CNN, P-CNN, R-CNN을 각각 사용한다.

### 2.1 Basic CNN

일반적인 CNN은 다층 퍼셉트론(Multi-Layer Perceptron, MLP) 기반의 발전된 형태이며 전처리 과정인 입력 이미지의 특징 추출 방식을 MLP에 결합한 신경망으로 최근 컴퓨터 비전, 오디오 프로세싱 등 다양한 분야에 적용되고 있다. CNN은 합성곱층(convolution layer)과 MLP로 구성되며 입력되는 합성곱층은 여러 개의  $m \times m$  크기의 커널 필터로 구성된 합성곱 필터와  $n \times n$  크기의 풀링층(pooling layer)으로 이루어진다. 입력되는 데이터의 구조는  $N \times N \times K$  크기의 2차원 이미지이며 커널 필터를 통해 입력이미

지의 특징 정보를 추출하고, 풀링층을 통해 입력된 이미지의 대푯값을 추출하여 차원을 감소시킴으로 과적합을 방지한다. 각 커널필터는 입력 이미지  $m \times m$  크기의 로컬 지역에 중첩되어 합성곱 연산을 수행하고, 이미지의 상단부에서 하단부로 이동하면서 중첩되는 각 로컬 지역에서 필터를 공유하여 동일한 연산을 수행함으로써 입력 이미지의 전체적인 특징값을 추출한다. 이때 커널 필터의 연산을 통해 출력된 특징맵은  $(N-m+1) \times (N-m+1)$  크기를 갖는다. 여러 종류의 풀링층 중 max-pooling filter는 입력 이미지의  $n \times n$  크기의 로컬 지역에서 최댓값을 대푯값으로 추출하고, 커널 필터와 마찬가지로 입력 이미지의 각 로컬 지역에서 동일한 과정을 수행하며  $(N-m+1) \times (N-m+1)$  크기의 입력 이미지를  $\frac{N-m+1}{n} \times \frac{N-m+1}{n}$  차원으로 감소시켜 특징맵을 추출한다. 각 합성곱층은 여러 개( $K$ )의 커널 필터와 max-pooling filter로 이루어져 있어  $K$ 개의 특징맵을 추출하고, 추출된 특징맵은 fully connected되어 MLP의 입력으로 사용한다. MLP는 은닉층과 softmax layer로 구성되어 입력 데이터 각 특징에 가중치를 부여하여 상위층으로 계산된 값을 전달하고, 최종 출력단인 softmax layer에서는 출력된 값들을 합하여 가장 높은 확률값을 갖는 레이블을 사운드 이벤트로 분류한다.

## 2.2 Residual CNN

딥러닝 기술이 발전함에 따라 효과적으로 입력데이터의 특징을 추출하기 위한 방식으로 여러 층으로 구성된 심층 신경망 방식이 제안되었다. 하지만 신경망의 층이 깊어짐에 따라 계산량의 증가뿐 아니라 gradient vanishing/exploding 문제를 초래할 수 있다. 따라서 이를 해결하기 위한 방법으로 각 층의 출력과 입력간의 차이를 학습시켜 기울기값 하락(degradation)문제를 해결하는 잔류학습 방식 기반의 R-CNN이 제안되고 있다.

Fig. 2는 잔류학습 방식을 적용한 R-CNN의 합성곱층 구조도이다. R-CNN은 여러 개의 잔류층(residual layer)으로 구성되고, 잔류층은 커널 필터, 활성화 함수인 ReLU(Rectified Linear Units), 데이터 정규화를 위

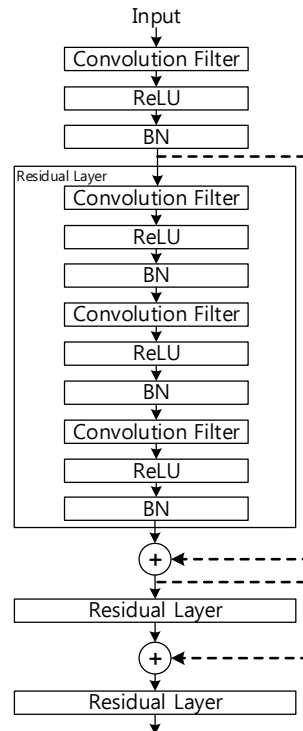


Fig. 2. Structure of the residual convolution layer.

한 배치 정규화(Batch Normalization, BN)로 구성된다. R-CNN은 특정 층의 입력값을 다음 층에 직접 전달하기 때문에 gradient vanishing 문제를 효과적으로 해결할 수 있고, 각 층의 입력값을 출력값에 더하는 identity mapping 방식을 사용하기 때문에 추가적인 파라미터와 복잡한 연산이 필요로 하지 않는 장점을 갖고 있다. 잔류 연결의 연산은 다음과 같이 나타낼 수 있다.

$$h_{k+1} = h_k + A(h_k, x_k; R_k), \quad (1)$$

여기서  $x_k$ ,  $h_k$ ,  $h_{k+1}$ 은 각각 입력 데이터,  $k$ 번째 유닛의 입력과 출력을 나타내며  $A$ 은 가중치 파라미터인  $R_k$ 를 이용한 잔류 함수이다. 따라서 잔류학습 방식은 gradient를 방해하지 않고 잔류연결로 구성된 레이어의 출력을 정제하여 학습 성능을 높이면서 깊은 심층망 구성을 가능하게 한다.

## 2.3 Parallel CNN

Fig. 3은 P-CNN의 전체 구조를 보여준다. P-CNN은

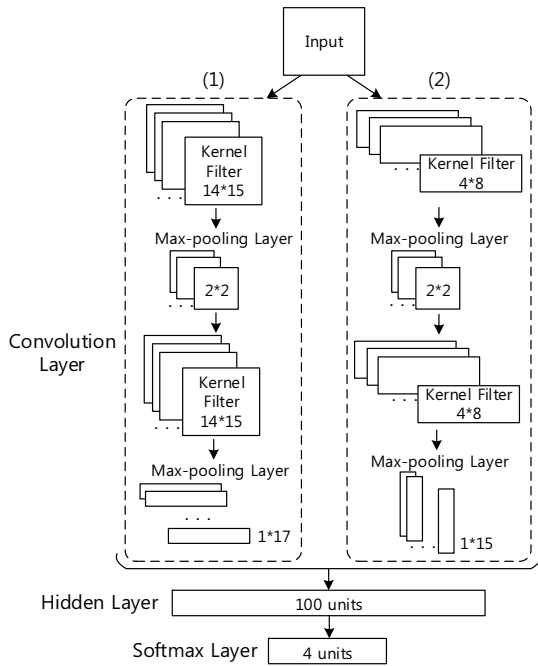


Fig. 3. Structure of the parallel convolutional neural network.

병렬로 구성된 두 개의 합성곱층과 직렬로 연결된 한 개의 MLP로 이루어진다. 각 합성곱층은 다른 크기의 커널 필터, max-pooling filter로 구성되어 필터 사이즈에 따른 다른 형태의 입력 데이터 특징을 각각 출력한다. 추출된 특징맵은 fully-connected되어 하나로 연결되고, MLP의 입력으로 사용된다. 본 논문에서는 병렬 구조의 합성곱층을 통해 오디오의 시간축, 주파수축 특징을 각각 추출하는 방식을 사용한다. 제한하는 P-CNN은 병렬 구조의 14×15, 4×8 크기의 커널 필터로 구성된 합성곱층, 추출된 특징맵의 차원 감소를 위한 2×2 크기의 max-pooling filter, 주파수 도메인과 시간축 도메인의 특징 추출을 위한 1×17, 15×1 크기의 max-pooling filter로 구성된다. Fig. 3의 Convolution Layer(1)의 구조는 주파수의 관계를 추출하기 위한 필터이며 추출된 특징맵은 주파수 축의 정보를 더 많이 보존하고 있다. 반면에 Fig. 3의 Convolution Layer(2)는 폐음의 시간 변화에 따른 관계를 추출하기 위한 필터로 이루어져 있으면 최종적으로 추출된 특징맵은 시간축 정보의 특징을 더 많이 보존한다. 각 합성곱층을 통해 추출된 특징맵은 fully connected되어 하나로 연결되고 MLP의 입력

으로 사용된다. 2.2에 서술된 MLP와 같이 P-CNN에서도 네트워크 모델의 최종 출력단인 softmax layer를 통해 출력된 값들을 합하여 가장 높은 확률값을 갖는 레이블을 사운드 이벤트로 분류한다.

### III. 실험 및 결과

#### 3.1 실험 데이터

본 논문에서는 건강한 폐음과 부잡음에 해당되는 수포음, 천명음, 협착음을 포함하여 네 가지 클래스를 선정하였으며 각 부잡음의 특징은 다음과 같다.

수포음은 연속성의 천명음과는 달리 폭발적인 단속성의 소리로, 주로 흡기 시에 ‘빠그락’하고 들리며 early inspiratory와 late inspiratory로 분류될 수 있다. Early inspiratory 수포음은 심한 심부전이나 만성 폐쇄성 폐질환에서 나타나고, late inspiratory 수포음은 세소기도의 폐쇄가 일어나는 모든 상황에서 발생하며 late inspiratory 수포음의 특수한 형태인 ‘velcro-like rale’는 간질성 폐질환에서 특징적으로 나타난다.

천명음은 좁아진 부위를 공기가 통과하면서 나는 ‘뽁’ 혹은 ‘휘’와 같은 소리이며 단조성과 복조성으로 분류될 수 있다. 단조성 천명음은 기도의 좁아진 부분이 호흡에 따라 변화하지 않을 때 생기며 기관지암이나 기도협착의 특징적인 소견을 갖고, 훨씬 흔한 복조성 천명음은 기도의 내강이 좁아져 있으나 그 정도가 균등하지 않을 때 나타나 만성 폐쇄성 폐질환에서 흔히 들을 수 있다.

협착음은 주로 숨을 들이쉴 때 중심 기도에서 크게 들리는 썩썩거리는 소리로 기도폐쇄를 시사하는 위험한 임상 징후를 판단하기 위한 소리이다.

본 연구의 폐음 데이터베이스는 건강한 피험자와 폐질환 환자의 폐음을 환자의 동의하에 녹음하여 사용하였다. 데이터베이스는 마이크를 부착한 청진기를 통해 피험자의 양와위 자세에서 흉부 후부의 다양한 기류 속도를 갖는 폐음을 녹음한 것이며 녹음하는 단계에서 저주파 대역의 잡음, 심장 박동 소리와 같은 외부 잡음은 차단 주파수가 80 Hz인 아날로그 저역 통과 필터를 적용하여 필터링하였다. 녹음

파일은 각 피험자마다 최대 흡기 기류로 가정하여 30 s 이내의 폐음으로 구성된다. 수집된 데이터 중 0.5 l/s 이상의 기류 값을 갖는 측정값을 임의로 선택하여 녹음된 신호의 기류 신호를 이용하여 전체 호흡주기를 결정하였으며 호흡시 흡기와 호기 구간을 구분하기 위해 차단주파수가 3 Hz인 저역 통과 필터를 이용하여 부드러운 기류 신호를 생성하였고, 기류 신호의 zero crossing positions에 따라 호흡의 시작 위치를 결정하였다. 전체 데이터베이스는 5명의 정상인과 23명의 환자로부터 수집한 5개의 건강한 폐음(138 s), 폐질환으로 인해 발생한 8개의 수포음(180 s), 11개의 천명음(150 s), 4개의 협착음(102 s) 파일로 구성되고, 각 녹음 파일은 16 kHz 샘플링레이트와 24 비트 해상도를 사용하였다. 학습 모델의 신뢰성과 데이터로부터의 영향을 최소화하기 위해 녹음된 오디오 신호를 프레임 단위로 분할한 뒤 각 클래스의 데이터를 랜덤하게 4개의 그룹으로 나누어 한 개의 그룹을 테스트 데이터로 사용하는 4-fold 교차 검증 방식을 이용하였다. 각 학습, 테스트 데이터 구성에 따른 4번의 모델 성능 측정 결과의 평균을 통해 폐음 분류 정확도를 획득하였다.

### 3.2 실험 환경 및 측정 방식

본 연구에서는 학습데이터의 25 %를 검증 데이터로 사용하여 그리드 서치를 통해 분류 정확도가 가장 높은 네트워크의 파라미터 값을 선택하였다. 실험에서 사용된 입력 데이터는 2 s에 해당하는  $98 \times 257$  크기의 스펙트로그램이고, 일반적인 CNN은  $5 \times 5$  크기의 64개 커널 필터와,  $3 \times 3$  크기의 max-pooling filter로 구성된 2개의 합성곱층, 이전과 동일한 크기의 128개 필터로 구성된 2개의 합성곱층을 사용하였다. P-CNN에서는 2.2.3에서 설명된 구조를 기반으로 각 32개의 필터를 사용하였고, R-CNN은 64개의 커널 필터, ReLU함수, batch normalization로 구성된 3개의 합성곱층으로 구성된 residual layer 6개를 사용하였다. 각 신경망의 MLP는 1개의 은닉층과 softmax layer로 구성된다. 학습의 최적화 방식으로 ADAM 알고리즘과 교차 엔트로피 오차를 사용하였다. 다양한 구조의 CNN 방식의 단채널 폐음의 분류 성능 비교를 위한 측정 지표로는 정밀도  $P$ 를 사용하

였고 계산식은 다음과 같다.

$$P = \frac{TP}{TP + FP}, \quad (2)$$

여기서 TP(True Positives)는 폐 질환으로 발생한 사운드 또는 건강한 폐의 소리가 해당하는 클래스로 제대로 분류된 경우를 의미하고, FP(False Positives)는 다른 클래스로 잘못 분류된 경우를 의미한다.

### 3.3 실험 결과

Table 1은 건강한 폐음과 부잡음을 이용한 폐질환 환자 식별 정확도를 보여준다. 실험 결과는 기존 방식인 k-NN과 CNN보다 잔류학습과 병렬 구조의 CNN을 사용했을 때 폐질환 환자 식별 성능이 더 우수함을 보여준다. Table 2 각각 다른 구조의 신경망을 이용한 4가지 클래스의 폐음 분류 성능 결과를 보여준다. k-NN은 폐음의 특징 패턴을 분석하여 비슷한 패턴을 군집화하는 방식으로 폐질환 환자 식별 성능에 비해 구체적 패턴 분석이 필요한 폐음 분류에서는 낮은 성능을 보이는 반면 일반적인 CNN은 k-NN에 비해 성능이 향상된 것을 확인할 수 있다. P-CNN의 폐음 분류 정확도는 95.36 %로 다른 방식의 신경망과 비교했을 때 가장 우수한 성능을 보이는 것을 확인할 수 있다. Table 3은 P-CNN기반의 분류 결과표와 클래스별 정확도를 보여준다. 수포음의 정밀도는

Table 1. The identification accuracy of patients with pulmonary diseases.

Method	Precision
k-NN	93.31 %
CNN	93.53 %
R-CNN	96.45 %
P-CNN	96.72 %

Table 2. Results of lung sound classification.

Method	Precision
k-NN	81.74 %
CNN	91.35 %
R-CNN	94.26 %
P-CNN	95.36 %

Table 3. Confusion matrix of P-CNN and precision.

		Predicted Class				Precision
		H	C	W	S	
Actual Class	H	21	0	1	1	91.30 %
	C	0	30	0	0	100 %
	W	1	0	24	0	96 %
	S	0	0	1	16	94.18 %

\*H: healthy, C: crackle, W: wheezing, S: stridor

100%로 가장 높은 성능으로 분류됨을 확인할 수 있다. 실험 결과를 통해 잔류학습을 적용한 심층 신경망 방식을 사용할 때 일반적인 CNN에 비해 폐음의 대표적인 특징을 보다 잘 추출하는 것을 확인할 수 있고, 병렬 구조의 합성곱층을 통해 여러 층을 쌓은 신경망 방식보다 시간 도메인, 주파수 도메인의 폐음 특징을 각각 추출하는 방식이 분류에 더 효과적임을 확인할 수 있다.

#### IV. 결 론

본 논문에서는 폐에서 발생하는 사운드를 통해 폐질환 환자를 식별하고, 부잡음을 분류하기 위해 다양한 CNN 방식을 이용하여 각 방식의 성능을 비교하였다. CNN 방식으로는 일반적인 CNN, P-CNN, R-CNN을 사용하였다. 실험을 통해 P-CNN 방식이 k-NN, 일반적인 CNN, R-CNN보다 폐음 분류에서 더 우수한 성능을 보이는 것을 확인하였다. 향후 제안된 방법을 기반으로 흉부 여러 위치에서 발생한 소리를 수집하여 다채널 폐음을 통한 구체적인 폐질환 식별을 위한 연구를 진행할 계획이다.

#### 감사의 글

본 논문은 2018년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2018R1D1A1B07041783). 그리고 2019년도 정부재원(과학기술정보통신부 여대학(원)생 공학연구팀제 지원사업)으로 과학기술정보통신부, 한국연구재단과 한국여성과학기술인지원센터의 지원을 받아 연구되었습니다.

#### References

1. R. Murphy, "Computerized multichannel lung sound analysis," IEEE Engineering in Medicine and Biology Magazine, **26**, 16 (2007).
2. S.-P. Yeo, C.-I. Jeon, S.-K. Yoo, D.-Y. Kim, and S.-H. Kim, "A study on robust pattern classification of lung sounds for diagnosis of pulmonary dysfunction in noise environment" (in Korean), Trans. of the Korea Institute of Electrical Eng. D, **51D**, 122-128 (2002).
3. N. Sengupta, M. Sahidullah, and G. Saha, "Lung sound classification using local binary pattern," arXiv:1710.01703 (2017).
4. A. Rizal, R. Hidayat, and H. A. Nugroho, "Lung sounds classification using spectrogram's first order statistics features," Proc. 6th Int. Annual Eng. Seminar (InAES) (2006).
5. D. Bardoua, K. Zhanga, and S. M. Ahmad, "Lung sounds classification using convolutional neural networks," Artificial Intelligence in Medicine, **88**, 58-69 (2018).
6. J. Pons, T. Lidy, and X. Serra, "Experimenting with musically motivated convolutional neural networks," Proc. 14th Int. Workshop on Content-Based Multimedia Indexing (2016).
7. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Proc. IEEE Conf. Computer Vision and Pattern Recognition, 770-778 (2016).

#### 저자 약력

▶ 김 지 연 (Gee Yeun Kim)



2018년 2월 : 광운대학교 전자융합공학과 학사  
2018년 3월 ~ 현재 : 광운대학교 전자융합공학과 석사과정

▶ 김 형 국 (Hyoung-Gook Kim)



1999년 ~ 2002년 : 독일 SIEMENS/Cortologic AG 책임연구원  
2002년 ~ 2005년 : 독일 베를린 공과대학교 Assistant Professor  
2005년 ~ 2007년 : 삼성종합기술원 수석연구원  
2007년 3월 ~ 현재 : 광운대학교 전자융합공학과 교수