

Parallel Network Model of Abnormal Respiratory Sound Classification with Stacking Ensemble

Myung-woo Nam*, Young-Jin Choi*, Hoe-Ryeon Choi*, Hong-Chul Lee*

*Student Researcher, Dept. of Industrial and Management Engineering, Korea University, Seoul, Korea

*Student Researcher, Dept. of Industrial and Management Engineering, Korea University, Seoul, Korea

*Visiting Professor, Dept. of Industrial and Management Engineering, Korea University, Seoul, Korea

*Professor, Dept. of Industrial and Management Engineering, Korea University, Seoul, Korea

[Abstract]

As the COVID-19 pandemic rapidly changes healthcare around the globe, the need for smart healthcare that allows for remote diagnosis is increasing. The current classification of respiratory diseases cost high and requires a face-to-face visit with a skilled medical professional, thus the pandemic significantly hinders monitoring and early diagnosis. Therefore, the ability to accurately classify and diagnose respiratory sound using deep learning-based AI models is essential to modern medicine as a remote alternative to the current stethoscope. In this study, we propose a deep learning-based respiratory sound classification model using data collected from medical experts. The sound data were preprocessed with BandPassFilter, and the relevant respiratory audio features were extracted with Log-Mel Spectrogram and Mel Frequency Cepstral Coefficient (MFCC). Subsequently, a Parallel CNN network model was trained on these two inputs using stacking ensemble techniques combined with various machine learning classifiers to efficiently classify and detect abnormal respiratory sounds with high accuracy. The model proposed in this paper classified abnormal respiratory sounds with an accuracy of 96.9%, which is approximately 6.1% higher than the classification accuracy of baseline model.

▶ **Key words:** Respiratory Sound Classification, Wheezes, Crackles, Convolutional Neural Network(CNN), Stacking Ensemble

-
- First Author: Myung-woo Nam, Corresponding Author: Hong-Chul Lee
 - *Myung-woo Nam (nmw0221@korea.ac.kr), Dept. of Industrial and Management Engineering, Korea University
 - *Young-Jin Choi (youngjin1206@korea.ac.kr), Dept. of Industrial and Management Engineering, Korea University
 - *Hoe-Ryeon Choi (annais@korea.ac.kr), Dept. of Industrial and Management Engineering, Korea University
 - *Hong-Chul Lee (hcllee@korea.ac.kr), Dept. of Industrial and Management Engineering, Korea University
 - Received: 2021. 09. 15, Revised: 2021. 11. 02, Accepted: 2021. 11. 11.

[요 약]

최근 코로나(Covid-19)의 영향으로 스마트 헬스케어 관련 산업과 비대면 방식의 원격 진단을 통한 질환 분류 예측 연구의 필요성이 증가하고 있다. 일반적으로 호흡기 질환의 진단은 비용이 많이 들고 숙련된 의료 전문가를 필요로 하여 현실적으로 조기 진단 및 모니터링에 한계가 있다. 따라서, 간단하고 편리한 청진기로부터 수집된 호흡음을 딥러닝 기반 모델을 활용하여 높은 정확도로 분류하고 조기 진단이 필요하다. 본 연구에서는 청진을 통해 수집된 폐음 데이터를 이용하여 이상 호흡음 분류모델을 제안한다. 데이터 전처리로는 대역통과필터(BandPassFilter)방법론을 적용하고 로그 멜 스펙트로그램(Log-Mel Spectrogram)과 Mel Frequency Cepstral Coefficient(MFCC)을 이용하여 폐음의 특징적인 정보를 추출하였다. 추출된 폐음의 특징에 대해서 효과적으로 분류할 수 있는 병렬 합성곱 신경망 네트워크(Parallel CNN network)모델을 제안하고 다양한 머신러닝 분류기(Classifiers)와 결합한 스택킹 앙상블(Stacking Ensemble) 방법론을 이용하여 이상 호흡음을 높은 정확도로 분류하였다. 본 논문에서 제안한 방법은 96.9%의 정확도로 이상 호흡음을 분류하였으며, 기본모델의 결과 대비 정확도가 약 6.1% 향상되었다.

▶ **주제어:** 호흡음 분류, 천명음, 수포음, 병렬 합성곱 신경망, 스택킹 앙상블

I. Introduction

전 세계적으로 호흡기 질환은 세계 사망 원인 중 세 번째로 높은 순위이며 천식, 폐암, 만성폐쇄성 질환, 급성 호흡기감염으로 매년 300만 명 이상의 사망자가 발생하고 있다 [1]. 호흡기 질환은 청진 또는 CT, X-ray 등 여러 방법을 통해 진단 가능하다. 특히, 청진은 최소 비용으로 호흡기 질환 검진 및 조기 진단에 있어서 가장 기본이 되며 중요한 진단방법이다[2]. 청진음을 통한 폐 질환의 진단은 보통 부잡음(adventitious sound)의 여부에 따라 구분된다. 부잡음은 폐의 이상 여부를 예측할 수 있는 대표적인 소리이며, 천명음(Wheeze)과 수포음(Crackle)로 나뉜다[3].

청진음을 통한 폐 질환의 진단은 숙련된 의료인이 반드시 필요하다. 현재 코로나의 확산으로 호흡기 관련 질환 발병 증가로 인하여 의료진 1명당 1,000명의 환자를 진료해야 하는 상황이며 앞으로 사상자 수는 지속적으로 증가할 것으로 추정된다[4]. 이에 청진음을 통해 호흡기 질환 환자에 대한 조기 진단 및 모니터링을 할 수 있는 전문 의료인력은 매우 부족하다. 따라서, 인공지능(artificial intelligence, AI)을 활용한 호흡기 질환을 자동으로 분석하는 진단 시스템은 의료 사각지대 해소에 큰 도움이 될 것으로 예상된다.

2017년에는 호흡음과 관련한 대규모 international conference on biomedical and health informatics (ICBHI) 공공데이터 셋이 공개됨에 따라 호흡음 분류에 관한 머신러닝 방법론과 딥러닝 방법론들의 많은 연구가

활발히 이루어졌다[5]. 초기 폐음 분류 연구에는 전통적인 머신러닝 방식을 통한 분류 연구에 초점이 맞추어져 있다. Jakovljević 등[6]의 연구에서는 hand-crafted 특징 추출과 은닉마코프(Hidden Markov) 모델을 이용한 호흡음 분류 방법론을 제안하였고, Serbes 등[7]의 연구에서는 머신러닝 서포트 벡터 머신(Support Vector Machine, SVM) 모델을 이용한 분류 방법론을 제안하였다. 최근에는 딥러닝이 기계학습 방식과 더불어 호흡음 분류 모델로서 많은 주목을 받고 있으며, 우수한 분류 성능을 보여주고 있다. 2차원의 스펙트로그램 이미지로 호흡음에 대한 특징 추출을 하고 합성곱 신경망(Convolutional Neural Network, CNN)[8]과 순환 신경망(Recurrent Neural Network, RNN)[9] 모델을 통하여 호흡음의 분류 성능을 비교하였다.

본 연구에서는 최적의 이상호흡음을 분류하기 위해 간단하면서도 효과적인 모델을 제안한다. 호흡기 알려지내과 전문의가 직접 수집한 데이터를 이용하여 효과적인 음성 데이터 특징 추출 기법인 로그 멜 스펙트로그램과 MFCC 방법을 통해 폐음의 특징적인 정보를 추출하고 딥러닝을 통해 이상호흡음을 분류하고자 한다. CNN 기반의 병렬네트워크 구조와 스택킹 앙상블 모델을 조합하여 분류 성능의 우수함을 보여주었다.

본 논문의 구성 및 내용은 다음과 같다. 제 2장에서는 호흡음 분류를 위한 전처리 방법론과 최근 딥러닝 네트워

크 알고리즘 연구에 관해 서술한다. 제 3장에서는 본 연구에 사용된 데이터 전처리 및 특징추출에 대해서 설명한다. 제 4장에서는 제안하는 모델 네트워크와 실험과 관련한 하이퍼파라미터에 대해 설명하고 제 5장에서는 사용된 평가 지표와 제안한 방법론의 성능에 대한 결과를 서술한다. 마지막 제 6장에서는 결론 및 향후 연구 방향을 제시한다.

II. Preliminaries

1. Related works

1.1 Parallel Input using CNN

Rocha 등[10]의 연구에서는 ICBHI 2017 데이터를 이용하여 스펙트로그램과 멜 스펙트로그램을 통해 특징을 추출하였으며, 두 개의 특징추출값을 병렬로 입력하여 CNN 모델을 제안하였다. 3개의 클래스(천명음, 수포음, 기타호흡음)와 2개의 클래스(천명음과 기타호흡음, 수포음과 기타호흡음)를 4개의 분류기(LDA, SVM, Boosted Trees, and CNN)를 통해 실험을 하였다. 3개의 분류 클래스로부터 96.9%의 가장 높은 분류 정확도를 보여주었다.

Minami 등[11]의 연구에서는 ICBHI 2017 데이터를 이용하여 스펙트로그램과 스칼로그램을 통해 특징추출 이후 2D 이미지로 변환하였다. 각 특징추출의 값을 VGG16 모델에 단일 또는 병렬로 입력한 모델의 성능을 비교하였을 때 병렬 입력으로 하는 예측 모델에서 성능이 우수하였다.

Ma 등[12]의 연구에서는 ICBHI 2017 데이터를 단시간 푸리에 변환(Short-Time Fourier Transform)과 웨이블릿(Wavelet) 분석을 통해 특징 추출을 하여 bilinear bi-ResNet 모델에 학습하여 네트워크를 제안하였고 4가지 클래스(정상, 수포음, 천명음, 수포음+천명음)에 대해서 50.16%의 성능을 얻었다.

1.2 Deep learning model with machine learning classifiers

Demir 등[13]의 연구에서는 폐음 신호를 스펙트로그램 2D 이미지로 변환하였으며 폐음 분류를 위해 2가지 딥러닝 모델 방식을 제안하였다. 먼저 VGG16 모델을 이용하여 특징추출 후 SVM 분류기를 사용하였고, 두 번째는 전이학습(Transfer learning)을 통해 VGG16 모델과 소프트맥스(Softmax) 분류기를 이용하여 4가지 호흡음(정상 호흡음, 천명음, 수포음, 그리고 천명음 플러스 수포음)을 분류하였다. VGG16과 SVM을 결합한 모델은 65.5%, 전이학습을 통한 방식은 63.09%의 정확도를 얻었다.

Wang 등[14]의 연구에서는 Chest X-ray 이미지 데이터셋을 사용하여 전이학습을 활용한 5가지 사전 학습된 딥러닝 모델(VGG16, InceptionV3, ResNet50, DenseNet121, and Xception)을 통하여 특징추출을 하였고 Xception 모델로부터 정상과 코로나 질병 2가지 클래스에 대해 정확도 96.75%의 성능을 얻었다. 성능 향상을 위해 추출된 특징들을 머신러닝 분류기(SVM, RandomForest, Decision, AdaBoost)를 통하여 이전 모델과 조합하여 비교실험을 하였다. Xception과 SVM을 결합한 모델에서 정확도 99.33%의 가장 높은 성능을 얻었다. Xception 단일모델과 비교하였을 때 약 2.58% 향상되었다.

Tran 등[15]의 연구에서는 drill 결점 탐지 시스템의 AI 기술을 접목한 연구로서 음성 신호로부터 멜 스펙트로그램과 스칼로그램을 이용하여 특징을 추출하고 이미지로 변환하여 이상 소리에 대해 3가지(Broken, Normal, Unrelated Sounds)로 분류를 하였다. VGG16 모델을 통한 특징값들을 추출하고 불필요한 특징값들을 NCA(Neighborhood Component Analysis)에 의해 차원 축소 하였다. 다양한 머신러닝 분류기와 조합하여 실험 예측한 결과 앙상블(Ensemble Subspace KNN) 모델이 최종적으로 80.25%의 분류성능을 보여주었다. 이전 선행 연구들에서는 두 개의 특징추출을 병렬로 연결한 모델과 사전 학습된 CNN 모델(VGG, ResNet)을 기반으로 여러 머신러닝 분류기를 통하여 비교 예측을 하였다.

본 연구에서는 선행연구보다 효과적인 CNN 모델을 제안하고 머신러닝 분류기를 구성하여 스택킹 앙상블 기법을 활용하였으며, 기존 모델보다 더 좋은 분류 성능을 보여주었다.

III. The Proposed Scheme

본 장에서는 호흡음 분류에 있어서 호흡 소리 데이터를 2개의 특징 추출을 통하여 병렬네트워크와 스택킹 앙상블 방법을 통해 호흡 이상호흡음을 효과적으로 분류할 수 있는 Fig. 1의 프레임워크를 제안하고자 한다.

1. Dataset

본 연구에 사용한 호흡음 데이터는 대학병원 알려지내과 전문의가 만 19세 이상 성인환자를 대상으로 총 31명(28명의 환자-3명의 정상)의 전 흉부 양측 상엽과 후 흉부 양측 하엽의 청진을 통해 데이터를 수집하였다. 연구에 사용한 데이터는 2020년 9월 IRB(Institutional Review Board) 심사 통과 후 2020년 11월부터 2021년

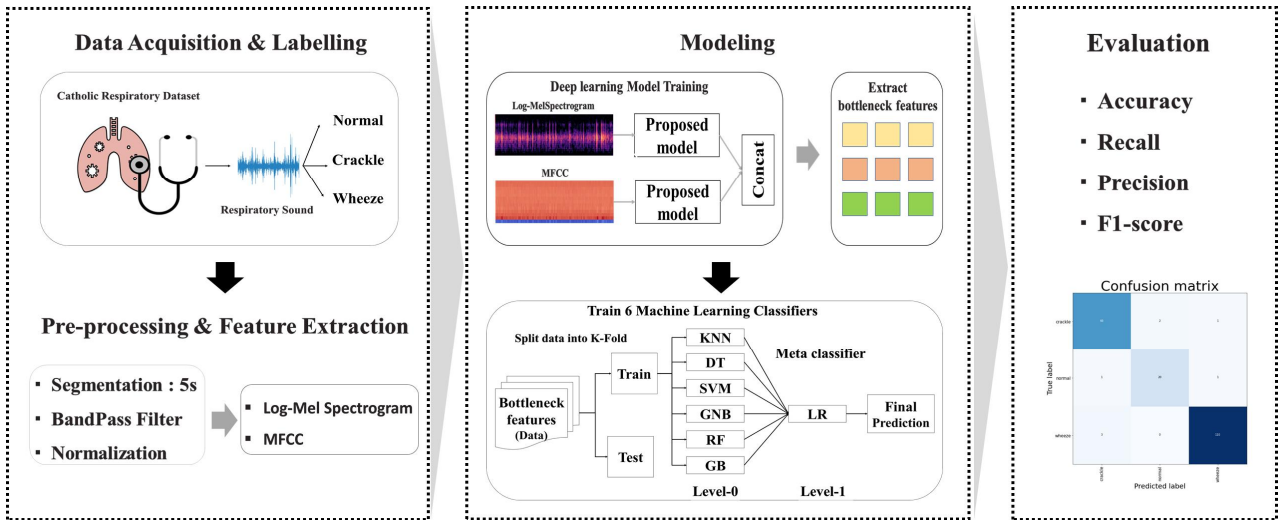


Fig. 1. Architecture of the framework for Respiratory Sound Classification

5월까지 호흡음을 수집하였다.

수집된 호흡음은 정상호흡음과 폐음 이상 징후를 알 수 있는 천명음, 수포음으로 구분되며 ‘정상호흡음’ 12개, ‘수포음’ 71개, ‘천명음’ 43개로 총 126개로 구성된다. 데이터는 전문의가 “3M Littmann” 디지털 청진기로 수집을 하였으며, 일부 데이터에는 잡음이 포함되어 있다. 클래스별 호흡음 데이터의 파형의 모습은 Fig. 2와 같다.

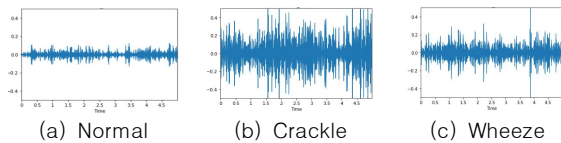


Fig. 2. Respiratory Sound (a) : Normal, (b) : Crackle, (c): Wheeze

2. Pre-Processing

소리 데이터는 서로 다른 sampling rate로 녹음되었을 경우 출력에 차이가 있다[2]. 호흡음의 sampling rate는 4,000Hz를 기준으로 했으며 데이터의 정규화(Normalization)와 호흡주기를 고려하여 5초 단위로 데이터를 단일화하였다.

모델을 학습 및 평가하기 위하여 전체 데이터를 8:2의 비율로 나누었으며, 이에 학습 데이터(Training data)는 816개, 평가 데이터(Testing data)는 205개로 구성되었다. Table 1은 전체 데이터에 대한 정보이다.

Table 1. Summary of the Training and Testing Sets

Class	Training set	Testing set	Total
Normal	91	23	114
Crackle	450	113	563
Wheeze	275	69	344
Total	816	205	1021

일부 데이터에 포함된 잡음을 제거하기 위해서 대역통과필터를 적용[16]하였으며 호흡음은 폭넓은 대역 범위가 존재하지만 수집된 데이터의 특성과 기존 호흡음 관련 연구에 적용된 대역대를 고려하여 250Hz ~ 1800Hz 주파수 대역대의 5th order Butterworth 대역통과필터를 사용하였다. 대역통과필터로부터 잡음이 제거된 음성은 다음 Fig. 3와 같다.

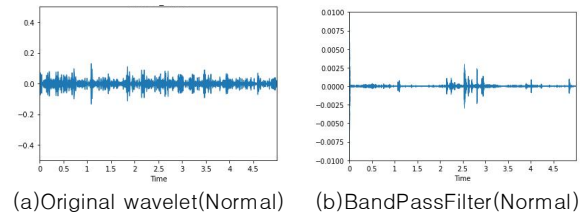


Fig. 3. Preprocessing of Respiratory Sound (a): Original wavelet(Normal), (b): BandPassFilter(Normal)

3. Feature Extraction

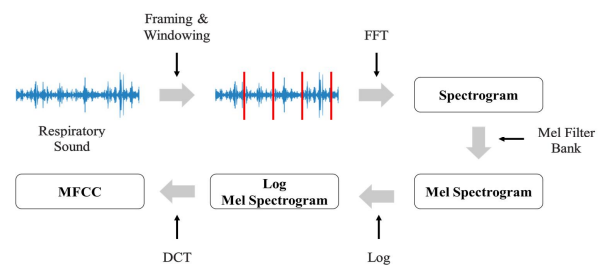


Fig. 4. Mel-scale frequency cepstral coefficient (MFCC) process of Respiratory sound

호흡음의 효과적인 특징을 추출하기 위해 로그 멜 스펙트로그램과 MFCC를 사용하였다. 음성 신호 분석 분야

에서 널리 사용되고 있는 방법론이며 음성 도메인 데이터를 시간-주파수 특징으로 추출하는 방식이다. 스펙트로그램은 오디오의 신호를 일정구간 분할하여 구간별 스펙트럼을 분석하여 특징을 추출하는 기법으로 음성신호의 시간의 변화에 따른 주파수의 세기를 파악하여 시각적으로 표현한 것이다[17]. 로그 멜 스펙트로그램과 MFCC 추출 과정은 다음과 같은 단계를 따른다.

먼저 호흡음 신호를 작은 크기의 프레임으로 자르며 시간 도메인을 주파수 성분으로 변환하는 고속 푸리에 변환(Fast Fourier Transform, FFT)를 적용하여 스펙트로그램을 얻는다. 그리고 저주파수 영역대에 주로 분포하는 호흡음의 특성을 고려하여 고주파수 대비 저주파수 영역을 더 세밀히 분석하는 멜 필터 बैं크(Mel Filter bank)를 통해 스펙트로그램을 멜 스펙트로그램으로 변환한다.

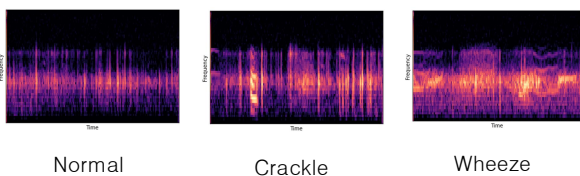
이후 멜 스펙트로그램에 로그 변환을 하여 본 연구에 사용한 첫 번째 음성 특징 벡터인 로그 멜 스펙트로그램(Log Mel Spectrogram)을 얻는다. 두 번째 음성 특징 벡터인 MFCC는 추출된 로그 멜 스펙트로그램에 이산 코사인 변환(Discrete Cosine Transform, DCT)을 통하여 추출한다. 전체적인 추출 과정에 대한 내용은 Fig. 4와 같다.

본 연구에서는 대역 통과 필터를 적용한 소리음원을 64ms(0.064x4,000=256)의 윈도우 길이로 75%씩 중첩하여 313개의 프레임을 구성하였다. 64개의 멜 밴드를 계산하여 64×313 크기의 멜 스펙트로그램을 추출하였다. 이후 로그를 취해 데시벨 값으로 변환하여 로그 멜 스펙트로그램을 얻었다(수식 1).

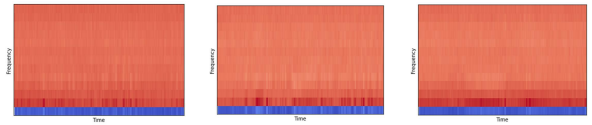
$$S_{dB} = 10 \log_{10} \left(\frac{S}{ref} \right) \quad (1)$$

MFCC는 로그 멜 스펙트로그램에 이산 코사인 변환을 적용하여 최종적으로 음성 특징 벡터 13×313 크기의 값을 얻었다.

본 연구에서는 서로 다른 두 개의 특징을 병렬방식으로 조합하여 사용하였다. 호흡음에 대하여 특징추출된 클래스에 대한 정보는 Fig. 5에서 보여주고 있다.



(a) Examples of Log-Mel spectrogram



(b) Examples of MFCC

Fig. 5. Respiratory Sound Feature Extraction (a): Mel spectrograms of respiratory sounds classes, (b): MFCC respiratory sounds classes

4. Classification Model

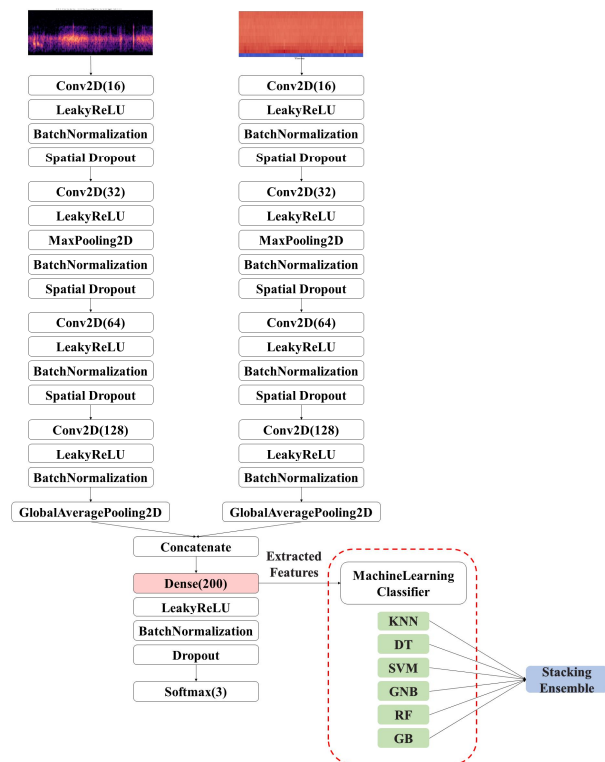


Fig. 6. The detailed architecture of proposed Parallel CNN model with stacking ensemble

4.1 Parallel CNN Model

본 연구에서 제시하는 분류 모델은 일반 심층 CNN 모델[10]에 SpatialDropout[18]을 적용하였으며, Flatten layer 대신 GlobalAveragePooling layer로 대체하는 등으로 모델을 재설계하였다. Fig. 6은 본 연구에서 제안한 병렬합성곱 모델의 구체적인 과정을 나타낸다. 특징추출은 두 개의 병렬 구조에서 ‘GlobalAveragePooling2D’ 레이어를 통해 학습하고 나온 특징들을 결합하여 마지막 Dense 층으로부터 200 차원의 특징값을 추출하였다. 이후 소프트맥스 분류기가 아닌 다양한 머신러닝 분류기를 조합하여 스택킹 앙상블을 통해 성능을 향상시켰다.

4.2 Stacking Ensemble Model

앙상블 기법은 여러 약한 학습기(Weak learner)를 조합하여 하나의 강한 학습기(Strong learner)를 생성함으로써 단일 모델에 비해 예측 오차를 줄이는 방법론이다. 앙상블 기법에는 대표적으로 데이터를 복원 추출하여 병렬적으로 학습한 이후 결과에 대한 투표(Voting) 방식인 배깅(Bagging) 기법[19], 배깅과 마찬가지로 데이터를 복원 추출하지만 순차적으로 학습하여 오답에 대해 높은 가중치를 부여하는 부스팅(Boosting) 기법[20], 마지막으로 단일 모델들의 예측 결과를 데이터로 학습하여 사용하는 스택킹 앙상블 기법이 있다. 본 연구에서는 단일모델의 단점을 보완할 수 있는 스택킹 기법을 사용하였다. 이는 여러 모형에서 예측된 값들을 다시 학습 데이터로 사용하고 최종 예측모형을 만드는 방법으로 각 알고리즘의 장단점을 보완하여 일반적으로 단일모델보다 성능이 좋은 것으로 알려져 있다[21].

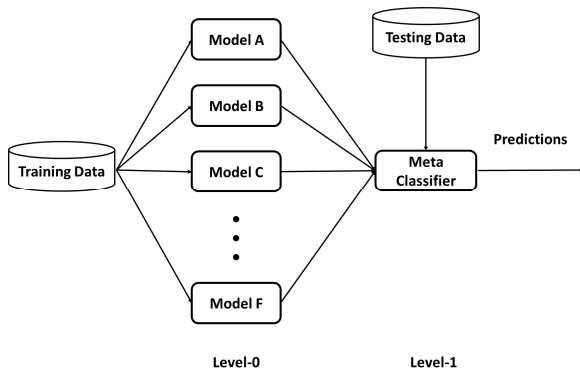


Fig. 7. Stacking Architecture

본 연구에서의 스택킹 앙상블 모델은 Fig. 7의 절차를 따른다. 먼저 Level-0에서는 제안한 모델로부터 나온 특징값들을 이용하여 여러 머신러닝 분류기의 예측 결과를 생성한다. 그 이후 Level-1에서는 Level-0에서 나온 결과를 학습 데이터로 사용하여 메타 분류기(Meta Classifier)를 생성한다. Level-1에서의 메타 분류기는 Level-0에서 사용한 학습 데이터와 다른 특성의 Data를 사용하므로 오버피팅을 방지하고 Bias를 줄이는 특성을 가진다. 스택킹 앙상블의 일반화 과정은 Level-0에서 사용하는 모델은 다양한 알고리즘을 사용하여 각 알고리즘으로부터 도출된 서로 다른 예측 결과값을 메타 분류기에서 입력 데이터로 활용할 수 있도록 하는 것이 좋다[22].

Table 2. Pseudo Code of the Stacking ensemble Algorithm

Algorithm Stacking
Input: training data $D = \{x_i, y_i\}_{i=1}^m$
Step 1. learn base-level classifiers for $t = 1$ to T do learn h_t based on D end for
Step 2. construct new data set of predictions for $i = 1$ to m do $D_h = \{x'_i, y_i\}$, where $x'_i = \{h_1(x_i), \dots, h_T(x_i)\}$ end for
Step 3. learn a meta-classifier learn H based on D_h return H
Output : ensemble classifier H

Table 2는 m개의 학습 Data를 x_i 와 y_i 로 구성하여 D로 표현한다. x_i 는 딥러닝 모델로부터 특징 추출되어 나온 벡터값을 의미하고 y_i 는 주어진 라벨값을 의미한다. Step 1에서는 학습데이터로 T개의 base-level의 분류기를 학습하여 h_t 분류 모형이 생성된다. Step 2에서는 이전 단계로부터 나온 예측값 x'_i 를 합쳐서 새로운 데이터 $D_h = \{x'_i, y_i\}$ 가 구성된다. Step 3에서는 새롭게 구성된 데이터 기반으로 메타 분류기를 학습하여 최종적인 예측값으로부터 스택킹 앙상블 분류기 H를 생성한다.

본 연구에서는 스택킹 앙상블 기법의 효율성을 높이고 모델들의 단점을 서로 보완할 수 있도록 K 최근접 이웃 알고리즘(K-NearestNeighbors), 의사결정나무, 서포트 벡터 머신, 가우시안 나이브 베이즈(Gaussian Naive Bayes), 랜덤포레스트와 그래디언트 부스팅(Gradient Boosting) 분류 알고리즘을 조합하여 사용하였다. 각 모델의 결과를 결합하는 메타 모델은 Simple Linear Model에 해당하는 로지스틱 회귀모델(Logistic Regression)을 사용하였다.

본 실험은 결과에 대한 신뢰성을 위해 클래스 비율을 고려한 StratifiedKFold 5겹 교차검증 방식을 이용하였다. 전체 데이터는 학습데이터와 평가 데이터로 8:2의 비율로 나누게 되며, 5번의 실험 결과의 평균을 구하여 최종적으로 호흡음 분류 성능을 측정하였다. 매 실험이 진행될 때 데이터가 무작위로 분할되기 때문에 이를 동일한 조건에서 비교하기 위해 난수를 같은 값으로 고정하였다.

IV. Experiments

1. HyperParameter Setting

본 연구 실험을 위해 손실함수는 교차 엔트로피(Cross Entropy)를 사용하였고, 최적화(Optimizer) 함수는 Adam으로 설정하였다. Adam Optimizer에 설정된 학습률(Learning rate)은 $1 \cdot e^{-4}$ 로 초기 설정을 하였고, 모델 학습이 일정 에폭(Epochs) 이상 개선이 없을 시 ReduceLROnplateau 함수를 통하여 학습률을 $1 \cdot e^{-5}$ 까지 줄여 학습을 안정적으로 수렴하게 하였다[23]. 훈련단계에서 오버피팅을 방지하기 위하여 Early Stopping을 적용하였다.

각 머신러닝 알고리즘 모형들과 최종 스택킹 앙상블 모델을 훈련할 때 설정되는 하이퍼파라미터(Hyper parameter)는 Table 3에 요약되어 있다.

Table 3. Optimum values of hyperparameters

Model	Hyper parameter	Value
K-Nearest Neighbors	n_neighbors	5
	function	Uniform
Decision Tree	min_samples_split	2
	min_samples_leaf	1
Support Vector Machine	C	1
	kernel	'rbf'
Gaussian Naive Bayes	var_smoothing	$1e^{-9}$
Random Forest	Estimators	100
	min_sample_split	2
	mins_sample_leaf	1
Gradient Boosting	Estimators	100
	subsample	1
	min_sample_split	2
	mins_sample_leaf	1
Logistic Regression	C	1
	max_iter	100

2. Evaluation

본 연구에서는 모델의 속성클래스 분류 성능을 평가하기 위해 혼동행렬(Confusion Matrix)을 사용하였다 (Table 4). 평가 척도는 정확도, 정밀도, 재현율, F1 점수를 사용하였고 수식 2-5로 계산된다. Table 4에서 행과 열은 실제 클래스와 예측 클래스로 분류한 속성을 나타낸다. N은 전체 평가 데이터의 관측치 개수를 의미한다.

Table 4. Confusion Matrix of Multiple Classification

		Predicted Class			
		Crackle	Wheeze	Normal	Total
Actual Class	Crackle	N_{11}	N_{12}	N_{13}	$\sum_{i=1}^3 N_{1i}$
	Wheeze	N_{21}	N_{22}	N_{23}	$\sum_{i=1}^3 N_{2i}$
	Normal	N_{31}	N_{32}	N_{33}	$\sum_{i=1}^3 N_{3i}$
	Total	$\sum_{i=1}^3 N_{i1}$	$\sum_{i=1}^3 N_{i2}$	$\sum_{i=1}^3 N_{i3}$	N

$$\text{Total Accuracy} = \frac{1}{N} \sum_{i=1}^3 N_{ii} \quad (2)$$

$$\text{Total Recall} = \frac{1}{3} \sum_{j=1}^3 \frac{N_{jj}}{\sum_{i=1}^3 N_{ij}} \quad (3)$$

$$\text{Total Precision} = \frac{1}{3} \sum_{j=1}^3 \frac{N_{jj}}{\sum_{i=1}^3 N_{i1}} \quad (4)$$

$$\text{Total F1-score} = \frac{2 \times (\text{Recall}_i \times \text{Precision}_i)}{\text{Precision}_i + \text{Recall}_i} \quad (5)$$

3. Experiment Result

Table 5는 각 방법론을 Table 4의 수식 2-5를 활용하여 평가한 결과이다. 본 연구에서 기본모델(Parallel CNN)과 제안모델(Parallel CNN Stacking Ensemble Model) 5개 Fold별 분류 결과를 보여준다. 기본모델에 비해 단일 머신러닝 분류기를 통하여 전반적으로 높은 성능을 보이고 있으며 그래디언트 부스팅 모델이 높은 정확도를 보여주고 있다. 다양한 머신러닝 분류기를 통하여 예측된 결과를 결합한 스택킹 앙상블 모델의 분류 결과는 2번째 Fold를 제외하고 모든 지표에서 가장 높은 성능을 얻었다.

Table 5. Classification performance in each fold

Models	Metric	Performance					
		Fold1	Fold2	Fold3	Fold4	Fold5	Avg
Baseline Model (Parallel CNN)	Accuracy	0.898	0.926	0.917	0.907	0.917	0.913
	F1-score	0.870	0.887	0.857	0.873	0.898	0.877
K-Nearest Neighbors	Accuracy	0.951	0.946	0.946	0.956	0.936	0.947
	F1-score	0.938	0.923	0.904	0.946	0.905	0.923
DecisionTree	Accuracy	0.917	0.931	0.936	0.912	0.922	0.924
	F1-score	0.906	0.892	0.898	0.872	0.847	0.883
Support Vector Machine	Accuracy	0.956	0.961	0.971	0.956	0.946	0.958
	F1-score	0.936	0.945	0.949	0.932	0.915	0.935
Gaussian Naive Bayes	Accuracy	0.902	0.833	0.848	0.858	0.848	0.857
	F1-score	0.874	0.788	0.788	0.802	0.795	0.809
Random Forest	Accuracy	0.951	0.971	0.971	0.956	0.931	0.956
	F1-score	0.921	0.947	0.938	0.937	0.878	0.924
Gradient Boosting	Accuracy	0.937	0.985	0.971	0.975	0.951	0.963
	F1-score	0.935	0.988	0.949	0.965	0.919	0.951
Parallel CNN Stacking Ensemble Model	Accuracy	0.961	0.980	0.980	0.975	0.951	0.969
	F1-score	0.942	0.975	0.963	0.966	0.930	0.955

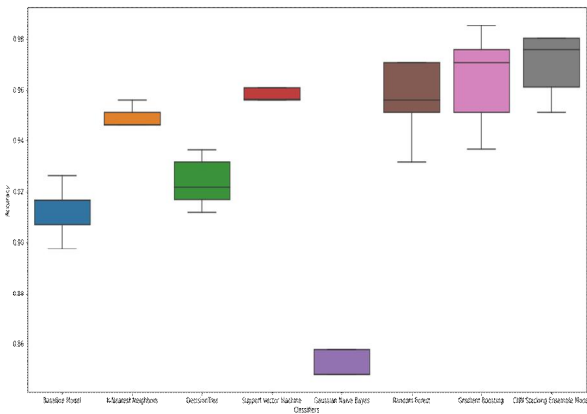


Fig. 8. Box plot analysis of accuracy among algorithms

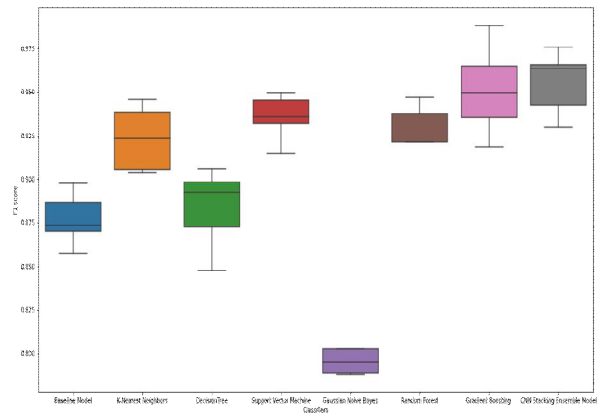


Fig. 9. Box plot analysis of F1-score among algorithms

Fig. 8과 9는 각 알고리즘에 따른 정확도와 F1 점수를 보여주는 박스플롯이다. 5겹 교차 검증을 통하여 실험을 진행하였고 단일 머신러닝을 이용한 모델은 기본모델보다 전반적으로 더 높은 성능을 보여주고 있다. 본 연구에서 제안한 단일 머신러닝 모델들은 알고리즘에 따라 서로 다른 성능을 보여주고 있으며 대체로 부스팅 모델에서 높은 성능 나타나고 있다. 최종적으로 본 연구에서 제안모델의 결과는 5번의 결과값(정확도, F1 점수)으로부터 평균값과 중앙값이 다른 모델에 비교하였을 때 가장 높게 나타난 것을 확인할 수 있다.

Table 6은 병렬 합성곱 신경망으로 분류한 기본모델과 제안모델의 결과를 각 Fold의 평균값으로 성능을 비교한 표이다. 제안모델의 정확도는 96.9%, 재현율은 94.6%, 정밀도는 97.4%, F1 점수는 95.5%로 기본모델 대비 증가율은 정확도는 약 6.1%, 재현율은 약 8.3%, 정밀도는 약 10.3%, F1 점수는 약 8.8%로 증가하였다.

Table 6. Experimental results(Average Classification Performance)

Models	Accuracy (±std)	Recall (±std)	Precision (±std)	F1 Score (±std)
Baseline Model (Parallel CNN)	91.3% (±0.010)	87.3% (±0.017)	88.3% (±0.016)	87.7 (±0.014)
Parallel CNN Stacking Ensemble Model	96.9% (±0.012)	94.6% (±0.028)	97.4% (±0.003)	95.5% (±0.017)
Improvement Ratio	6.1%	8.3%	10.3%	8.8%

Fig. 10은 3가지 호흡음 이상 유무의 실제값과 예측값에 대한 혼동행렬을 보여준다. (a)는 기본모델의 분류 결과이며 (b)는 제안모델의 분류 결과이다. Fig. 10에서 가로축은 실제 라벨값을 의미하며 세로축은 모델이 산출한 예측값을 의미한다. 예측값이 높으면 높을수록 이상호흡음에 대한 클래스에 대해서 정확하게 예측하였다고 해석할 수 있다. 따라서, 기본모델의 예측결과는 수포음 클래스

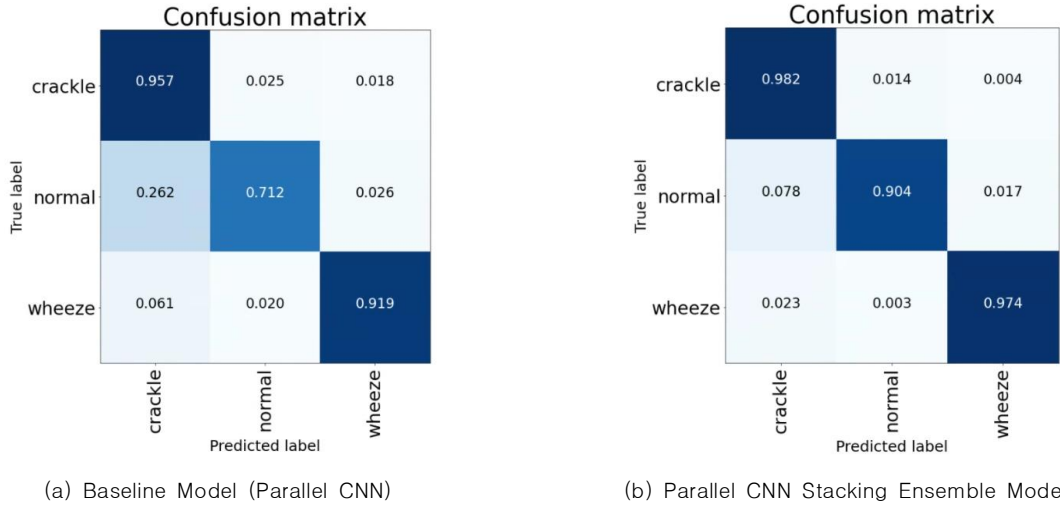


Fig. 10. Confusion matrix for the ensemble respiratory sound classification models :

(a) (Baseline) Parallel CNN model analysis with Softmax Classifier; (b) Parallel CNN with Stacking Ensemble analysis

스와 천명음 클래스에서 모델이 다소 오답으로 예측분류하였다. 하지만, 제안모델의 결과에서 기존과 비교하였을 때 수포음과 정상 클래스에서 분류 결과가 향상된 것을 볼 수 있다.

4. Statistical Test

본 연구는 기본모델과 제안모델의 분류 성능이 통계적으로 유의미한 차이가 있는지 확인하였다. 이표본 t 검정 (Two Sample t-test)을 수행한 결과는 Table 7과 같다. 검정 결과 p-value가 0.05보다 작고 95%의 신뢰구간 안에 0을 포함하지 않아 두 모델의 분류 성능이 통계적으로 유의미한 차이가 있는 것을 확인하였다.

Table 7. Two sample t-test experimented in this study

Two Sample t-test		
Hypothesis	p-value	95% C.I.
$H_0: \mu_{base} = \mu_{proposed}$	p-value < 0.05	[0.03, 0.06]
$H_1: \mu_{base} \neq \mu_{proposed}$		

V. Conclusions

본 연구에서는 호흡기 알려지내과 전문의가 제공한 호흡음 데이터를 사용하여 호흡음에 적합한 전처리 방식과 특징추출을 통하여 이상 호흡음에 대해 효과적으로 분류할 수 있는 병렬 합성곱 신경망 모델을 제안하였다. 소프트맥스 분류기를 사용한 Baseline의 결과와 다양한 머신러닝 분류기를 사용한 결과를 비교하였으며, 단일 머신러닝 분류기를 사용하였을 때 전반적으로 높은 성능을 보이

는 것을 확인할 수 있었다. 또한 다양한 머신러닝 분류기를 통해 도출된 결과를 결합하여 사용하는 스택킹 앙상블 기법이 모든 성능지표에서 가장 우수한 결과를 나타내는 것을 확인할 수 있었다. 본 연구는 수집된 환자들의 호흡음 수의 부족과 세분화된 호흡질환분류의 한계점이 있으므로, 추가로 데이터 증강기법 적용과 호흡관련 질환분류 모델을 위한 연구가 필요하다.

본 연구를 통해 향후 인공지능을 접목한 기술을 통해 폐질환 이상여부를 조기에 스크리닝하여 신속한 치료가 가능해질 수 있으며 사회적 취약계층들에게 의료 접근 방식 개선 및 의료 서비스의 질 향상에 도움이 될 것으로 기대 된다.

ACKNOWLEDGEMENT

This work was supported by Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No. 2020-0- 02199, “For building telemedicine environment, AI -based cardiovascular and Lung disease classification model development.”) and supported by Brain Korea 21 FOUR.

REFERENCES

- [1] WHO, The global impact of respiratory disease. European Respiratory Society, 2017.
- [2] A. Bohadana, G. Izbicki, and S. S. Kraman, "Fundamentals of lung auscultation," *N. Engl. J. Med.*, vol. 370, no. 8, pp. 744-751, 2014.
- [3] P. S. Faustino, "Crackle and wheeze detection in lung sound signals using convolutional neural networks," 2019.
- [4] J. Xie, Z. Tong, X. Guan, B. Du, H. Qiu, and A. S. Slutsky, "Critical care crisis and some recommendations during the COVID-19 epidemic in China," *Intensive Care Med.*, vol. 46, no. 5, pp. 837-840, 2020.
- [5] B. M. Rocha et al., "An open access database for the evaluation of respiratory sound classification algorithms," *Physiol. Meas.*, vol. 40, no. 3, p. 35001, 2019.
- [6] N. Jakovljević and T. Lončar-Turukalo, "Hidden markov model based respiratory sound classification," in *International Conference on Biomedical and Health Informatics*, 2017, pp. 39-43.
- [7] G. Serbes, S. Ulukaya, and Y. P. Kahya, "An automated lung sound preprocessing and classification system based on spectral analysis methods," in *International Conference on Biomedical and Health Informatics*, 2017, pp. 45-49.
- [8] M. Aykanat, Ö. Kılıç, B. Kurt, and S. Saryal, "Classification of lung sounds using convolutional neural networks," *EURASIP J. Image Video Process.*, vol. 2017, no. 1, pp. 1-9, 2017.
- [9] D. Perna and A. Tagarelli, "Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks," in *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, 2019, pp. 50-55.
- [10] B. M. Rocha, D. Pessoa, A. Marques, P. Carvalho, and R. P. Paiva, "Automatic Classification of Adventitious Respiratory Sounds: A (Un) Solved Problem?," *Sensors*, vol. 21, no. 1, p. 57, 2021.
- [11] K. Minami, H. Lu, H. Kim, S. Mabu, Y. Hirano, and S. Kido, "Automatic classification of large-scale respiratory sound dataset based on convolutional neural network," in *2019 19th International Conference on Control, Automation and Systems (ICCAS)*, 2019, pp. 804-807.
- [12] Y. Ma et al., "LungBRN: A smart digital stethoscope for detecting respiratory disease using bi-resnet deep learning algorithm," in *2019 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2019, pp. 1-4.
- [13] F. Demir, A. Sengur, and V. Bajaj, "Convolutional neural networks based efficient approach for classification of lung diseases," *Heal. Inf. Sci. Syst.*, vol. 8, no. 1, pp. 1-8, 2020.
- [14] D. Wang, J. Mo, G. Zhou, L. Xu, and Y. Liu, "An efficient mixture of deep and machine learning models for COVID-19 diagnosis in chest X-ray images," *PLoS One*, vol. 15, no. 11, p. e0242535, 2020.
- [15] T. Tran and J. Lundgren, "Drill Fault Diagnosis Based on the Scalogram and Mel Spectrogram of Sound Signals Using Artificial Intelligence," *IEEE Access*, vol. 8, pp. 203655-203666, 2020.
- [16] Y. Ma, X. Xu, and Y. Li, "LungRN+ NL: An Improved Adventitious Lung Sound Classification Using Non-Local Block ResNet Neural Network with Mixup Data Augmentation," in *Interspeech*, 2020, pp. 2902-2906.
- [17] A. M. Badshah, J. Ahmad, N. Rahim, and S. W. Baik, "Speech emotion recognition from spectrograms with deep convolutional neural network," in *2017 international conference on platform technology and service (PlatCon)*, 2017, pp. 1-5.
- [18] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler, "Efficient object localization using convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 648-656.
- [19] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123-140, 1996.
- [20] R. E. Schapire, "The strength of weak learnability," *Mach. Learn.*, vol. 5, no. 2, pp. 197-227, 1990.
- [21] Y. Xiao, J. Wu, Z. Lin, and X. Zhao, "A deep learning-based multi-model ensemble method for cancer prediction," *Comput. Methods Programs Biomed.*, vol. 153, pp. 1-9, 2018.
- [22] B. Zhai and J. Chen, "Development of a stacked ensemble model for forecasting and analyzing daily average PM_{2.5} concentrations in Beijing, China," *Sci. Total Environ.*, vol. 635, pp. 644-658, 2018.
- [23] T. Gabruseva, D. Poplavskiy, and A. Kalinin, "Deep learning for automatic pneumonia detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 350-351.

Authors



Myung-woo Nam received the B.S. degrees in statistics from University of California, Los Angeles, U.S. in 2018 and M.S. in Industrial Engineering From Korea University, Korea in 2022 respectively.

Myung-woo Nam is interested in Big data analysis, Machine Learning and Deep Learning.



Young-Jin Choi received the B.S. degree in Biomedical Science from Daegu University, Korea, in 2015. He is currently a Integrated M.S/Ph.D Student in the department of Industrial Management Engineering at Korea

University. He is interested in Artificial Intelligence, Health Care and Simulation.



Hoe-Ryeon Choi received the B.S. degree in Industrial Engineering from Dankook University, Korea in 1993. She finished the doctor coursework in Industrial Engineering at Korea University, Korea in 2005 respectively.

Dr. Choi currently working as a visiting professor in School of Industrial and Management Engineering at Korea university. She is interested in Big Data analysis, Machine Learning and AI Algorithms(Image Generation, Autonomous Driving)



Hong-Chul Lee received the B.S. degree in Industrial Engineering from Korea University, Korea, in 1983, M.S. degree in Industrial Engineering from Texas Arlington University, U.S. in 1988 and he received

Ph.D. degree in Industrial Engineering from Texas A&M University, U.S. respectively. Dr. Lee joined the faculty of the Department of Industrial Management Engineering at Korea University, Seoul, Korea, in 1996. He is currently a Professor in the Department of Industrial Management Engineering, Korea University. He is interested in Artificial Intelligence, Manufacturing Engineering System, and Simulation.