IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 7, 2022, accepted May 8, 2022, date of publication May 12, 2022, date of current version May 23, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3174678

Lightweight Skip Connections With Efficient Feature Stacking for Respiratory Sound Classification

YOUNGJIN CHOI[1], HOERYEON CHOI[1], HWAYOUNG LEE[2], SOOKYOUNG LEE[2], AND HONGCHUL LEE[2]

[1]School of Industrial Management Engineering, Korea University, Seoul 02841, South Korea
[2]Division of Allergy, Department of Internal Medicine, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul 06591, Republic of Korea

Corresponding author: Hongchul Lee (hclee@korea.ac.kr)

This work was supported in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant funded by the Korean Government through Ministry of Science and ICT (MSIT) (for building telemedicine environment, AI-based cardiovascular and lung disease classification model development) under Grant 2020-0-02199, and in part by the Brain Korea 21 FOUR.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Clinical Research Ethics Committee of the Catholic Medical Center under Application No. KC20ONSI0774.

ABSTRACT As the number of deaths from respiratory diseases due to COVID-19 and infectious diseases increases, early diagnosis is necessary. In general, the diagnosis of diseases is based on imaging devices (e.g., computed tomography and magnetic resonance imaging) as well as the patient's underlying disease information. However, these examinations are time-consuming, incur considerable costs, and in a situation like the ongoing pandemic, face-to-face examinations are difficult to conduct. Therefore, we propose a lung disease classification model based on deep learning using non-contact auscultation. In this study, two respiratory specialists collected normal respiratory sounds and five types of abnormal sounds associated with lung disease, including those associated with four lung lesions in the left and right anterior chest and left and right posterior chest. For preprocessing and feature extraction, the noise was removed using three pass filters (low, band, and high), and respiratory sound features were extracted using the Log-Mel Spectrogram-Mel Frequency Cepstral Coefficient followed by feature stacking. Then, we propose a lung disease classification model of dense lightweight convolutional neural network-bidirectional gated recurrent unit skip connections using depthwise separable convolution based on the extracted respiratory sound information. The performance of the classification model was compared with both the baseline and the lightweight models. The results indicate that the proposed model achieves high performance and has an accuracy of 92.3%, sensitivity of 92.1%, specificity of 98.5%, and f1-score of 91.9%. Using the proposed model, we aim to contribute to the early detection of diseases during the COVID-19 pandemic.

INDEX TERMS Densely BiGRU connection, lightweight convolutional neural network, log-mel spectrogram-mel frequency cepstral coefficients, pass filter feature stacking, respiratory sound.

I. INTRODUCTION

According to the World Health Organization, chronic obstructive pulmonary disease (COPD) was the third leading cause of death worldwide in 2020. In addition, according to the 2017 report from the International Respiratory Society, approximately 333 million people had asthma in 2014 [1]. Similarly, pneumonia is one of the three leading causes of death and disability in both children and adults [2]. Therefore, given the high social and economic costs associated with lung disease, routine monitoring for early diagnosis of the disease has attracted considerable research focus. Due to the recent surge in COVID-19, lung disease is diagnosed using imaging equipment such as pulmonary function tests, X-rays, computed tomography (CT), and magnetic resonance imaging (MRI). However, because this equipment is expensive, it is a limitation of this technology that it requires the presence of skilled medical staff depending on

The associate editor coordinating the review of this manuscript and approving it for publication was Stavros Ntalampiras.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

the patient [3], [4]. Therefore, it is necessary to study artificial intelligence (AI) using a stethoscope that can diagnose early and solve the problems of the shortage of beds and the rapidly increasing number of critically ill patients [5].

Auscultation is a technology that allows medical staff to detect lung sounds and identify lung diseases using a simple and inexpensive non-invasive method using a stethoscope [6]. Accordingly, studies have been conducted to classify lung diseases. Auscultation of the chest is the most efficient method for measuring lung dysfunction. It enables early diagnosis and directly provides characteristic changes in respiration, which reduces the risk of exposure to X-ray radiation and infectious diseases during a pandemic [7], [8].

In the lung disease-related research, diseases were classified using machine learning and deep learning techniques through X-rays, COVID-19, and respiratory sound datasets. Kc *et al.* [9] classify diseases of healthy, bacterial pneumonia, COVID-19, and viral pneumonia using ImageNet, which is pre-trained for chest X-ray images. Despotovic *et al.* [10] constructed respiratory, negative, and cough datasets for COVID-19 positive and negative patients. They extracted features through vggish and identified COVID-19 using random forest, bagging, boosting, and multi-layer perceptron (MLP) during machine learning.

In addition, with the rapid spread of telemedicine services during an era of remote work, for patients, the medical system is changing from a visit treatment method to a preventive health management method that uses big medical data. With the development of mobile technology, diseases are diagnosed using convenient and straightforward wearable-based small mobile medical devices at home. This modification enables remote diagnosis by storing the patient's biometric and medical information in the device and transmitting it to medical staff [11]. However, AI requires a high-performance central processing unit (CPU) and graphic processing unit (GPU), so a medical stethoscope capable of diagnosing a disease equipped with artificial intelligence is insufficient. Because it is part of the actual patient diagnosis, the accuracy is high, and the system must be time and memory-efficient for use in real-time applications [12]. Accordingly, the latest AI technology is being applied to computer network healthcare systems such as the internet of things (IoT) [13] cloud and edge computing technologies. Alotaibi and Subahi [14] proposed a method of deriving medical business goals using goal-oriented requirements extraction approach (GOREA), which includes modeling e-health business requirements and e-health IT and system requirements. For system verification, patients were identified and diagnosed through essential operations and services of the hospital emergency room for COVID-19 patients.

Therefore, we propose an artificial intelligence model for lung disease using auscultation in this study. We performed objective and accurate disease prediction and diagnosis through an artificial intelligence model using respiratory sounds collected from patients suffering from lung disease. It is a model that applied staking feature extraction, combined depthwise separable convolution, and skip connections based on lightweight with few parameters. The model showed superiority in high classification performance and fast reasoning time, and it enables efficient and stable decision-making by emergency patients and medical staff using breathing sounds collected from a stethoscope [15].

The remainder of this paper is organized as follows: Section II. discusses the classification of lung diseases and previous studies related to weight reduction. Section III. describes the data, the convolutional neural network (CNN)-bidirectional gated recurrent unit (BiGRU) connections model proposed in this study, and the method used to denoise and extract the features of respiratory sounds. Section IV. introduces the experimental environment and evaluation indicators used in the study, and Section V. presents the experimental results. Finally, Section VI. describes the discussion and Section VII. presents conclusions, limitations, and future research directions.

II. RELATED WORKS

A. DEEP NEURAL NETWORK ARCHITECTURES

Hsu *et al.* [16] presented a deep learning-based classification model, the CNN-BiGRU, which uses stridor and rhonchi measured using a 3M Littmann 3200 stethoscope. Shi *et al.* [17] classified asthma, pneumonia, and normal diseases using a 3M Littmann 3200 stethoscope. A model combining the visual geometry group (VGG) model and BiGRU was proposed for the effective classification of lung disease, and the model demonstrated an accuracy of 87%. Acharya and Basu [18] proposed a CNN-RNN hybrid model after feature extraction that employs a Mel spectrogram using the international conference in biomedical and health informatics (ICBHI) dataset. This model is characterized by bidirectional long short-term memory (BiLSTM), and it demonstrated sensitivity of 48% and specificity of 84%. Cakır *et al.* [19] proposed a convolutional recurrent neural network (CRNN) classification model for the TUT Sound Event dataset sounds. They verified the advantages of spectrogram feature extraction and CRNN audio signals. Peng *et al.* [20] presented a model for classifying sounds from the Urban Sound 8k datasets (e.g., air conditioner, car horn, children playing, dog bark drill, engine idling, gunshot, jackhammer, siren, and street music). A classification model using only a gated recurrent unit (GRU) was proposed for the Mel spectrogram - mel frequency cepstral coefficient (MFCC) and Log-Mel spectrogram-MFCC, and an accuracy of 94.3% was obtained.

Li *et al.* [6] proposed a deep learning architecture LungAttn that integrates Augmented Attention Convolution with ResNet block to improve the accuracy of lung sound classification using the ICBHI dataset. After resolving data imbalance through a mix-up, preprocessing using pass filter to maintain the frequency band of 50-2000Hz, extracting features through triple short-time fourier transform (STFT) Q-factor wavelet transform. The model's performance improved by 1.69% compared to the latest model of

ICBHI, confirming the effect of lung sound classification. Jayalakshmy and Sudha [21] presented a respiratory sound classification model using a pre-trained optimized Alexnet. Using empirical mode decomposition (EMD), respiratory sound characteristics were extracted using a scalogram of the respiratory sound signals, which were divided into several intrinsic mode functions (IMF). The model demonstrated improved performance of 83.78% compared to the existing wavelet method. Demir *et al.* [22] applied the spectrogram to the ICBHI set to extract respiratory sound information and proposed a parallel CNN model. As a characteristic of the study, the paper presented a model that merges linear discriminant analysis (LDA) and support vector machine (SVM) classifiers in a parallel network. Additionally, Demir *et al.* [23] extracted the bottleneck feature from the spectrogram after STFT transformation using the VGG16 model and obtained a performance of 65.5% by using the SVM classifier for the feature.

In the study using a 1D CNN and RNN, Basu *et al.* [24] extracted the MFCC using the ICBHI Dataset, reduced it to one-dimensional data, and proposed a lung disease classification model that uses a GRU. As a result, a high classification performance of 95.6% was obtained. Lella and Pja [25] extracted features through a denoising autoencoder (DAE) using the COVID-19 respiratory sound set and proposed a respiratory 1D CNN classification model that obtained an accuracy of approximately 90%. However, 2D-based feature extraction requires different strategies.

B. DENSE SKIP CONNECTIONS

AI-based deep learning uses skip connections or residuals to prevent the overfitting of models in which learning is not performed correctly due to gradient loss as the neural network deepens in the learning process. Skip connections transfer valuable information between the layers of all feature maps, and it is easy to learn by reducing training time and improving convergence speed [26], [27]. Godin *et al.* [28] proposed skip connections model that improves the overfitting that occurs in the RNN layer stack in the language model, and they demonstrated the model's efficiency and excellence. Ullah *et al.* [29] presented a classification model that applied a 3D CNN and skip connections using 3D-based video data. Roy *et al.* [30] obtained an accuracy of 85% using a convolutional densely connected gated recurrent neural (ChronoNet) classification model using 1D-based TUH Abnormal electroencephalogram (EEG) corpus data. This model has the advantage of learning by densely connecting deeper layers by alleviating the problem of reducing the accuracy of training using densely connected. Shuvo *et al.* [12] presented a lightweight classification model of CNL-UNet using biomedical images of chest X-ray, dermatoscopy, microscopy, ultrasound, and MRI. This model uses the pre-trained weights of the VGG-16 model and applies the skip connections of Res-path to convolution for compatibility of encoder and decoder. As a result, overfitting was prevented, and learning efficiency was improved by reducing the weight through a small number of parameters. The dice score of 95.94% and the jaccard index of 92.26% showed the model's superiority.

C. LIGHTWEIGHT MODELS

For use in embedded devices, Joshi *et al.* [13] attached a sensor to the patient's upper body to detect cough sounds and provided information to clinicians and caregivers using Thing Speak IoT-cloud technology. Cough sounds are automatically detected for respiratory diseases using a reduced CNN model. Accordingly, reducing the weight can be divided into reducing the size of a learning model or a method of efficiently designing a network using MobileNet's depthwise separable convolution. The research related to weight reduction suggests changing the model's input size or minimizing its structure. Shuvo *et al.* [31] proposed a lightweight model using filters of sizes 64, 64, 96, and 96 after feature extraction of ICBHI breath sounds using empirical mode decomposition (EMD) and continuous wavelet transform (CWT). The disease classification results showed an accuracy of greater than 98%. Jung *et al.* [32] proposed a respiratory sound classification model using a depthwise separable convolution CNN that fuses STFT and the MFCC.

Ponomarchuk *et al.* [33] proposed a machine learning method to detect COVID-19 based on breath, voice rapidly, and cough of open, private datasets with verified labels from hospitals and app data. They proposed a hybrid ensemble model that combines vggish and cochleagram, Mel spectrogram, and gradient boosting with patient information as feature values, lightweight CNN, and logistic regression. Therefore, the model was loaded in the mobile application and showed high performance on the public dataset and the collected noise data. Li *et al.* [34] presented a lightweight model that classifies the presence or absence of abnormalities in heart sounds, a characteristic similar to lung sounds. This lightweight model reduces the number of parameters based on a 2D CNN with only three convolution layers. The learning parameters decreased by approximately 90%, but the performance also decreased. In addition, there is insufficient prior research regarding specific lung disease classification and weight reduction focusing on lung symptoms (normal, wheeze, crackle). Therefore, in this study, the model was improved based on previous studies, and we introduce three special features:

- Respiratory specialists collected respiratory sounds from four lung lesions in the anterior and posterior thoracic regions and labeled them for disease information.
- Feature stacking was applied to diversify the bandwidth to improve the extraction of respiratory sound features.
- A lightweight model using depthwise separable convolution and BiGRU skip connections contribute to the early diagnosis and monitoring of diseases through high-performance artificial intelligence solutions in preparation for infectious respiratory diseases.

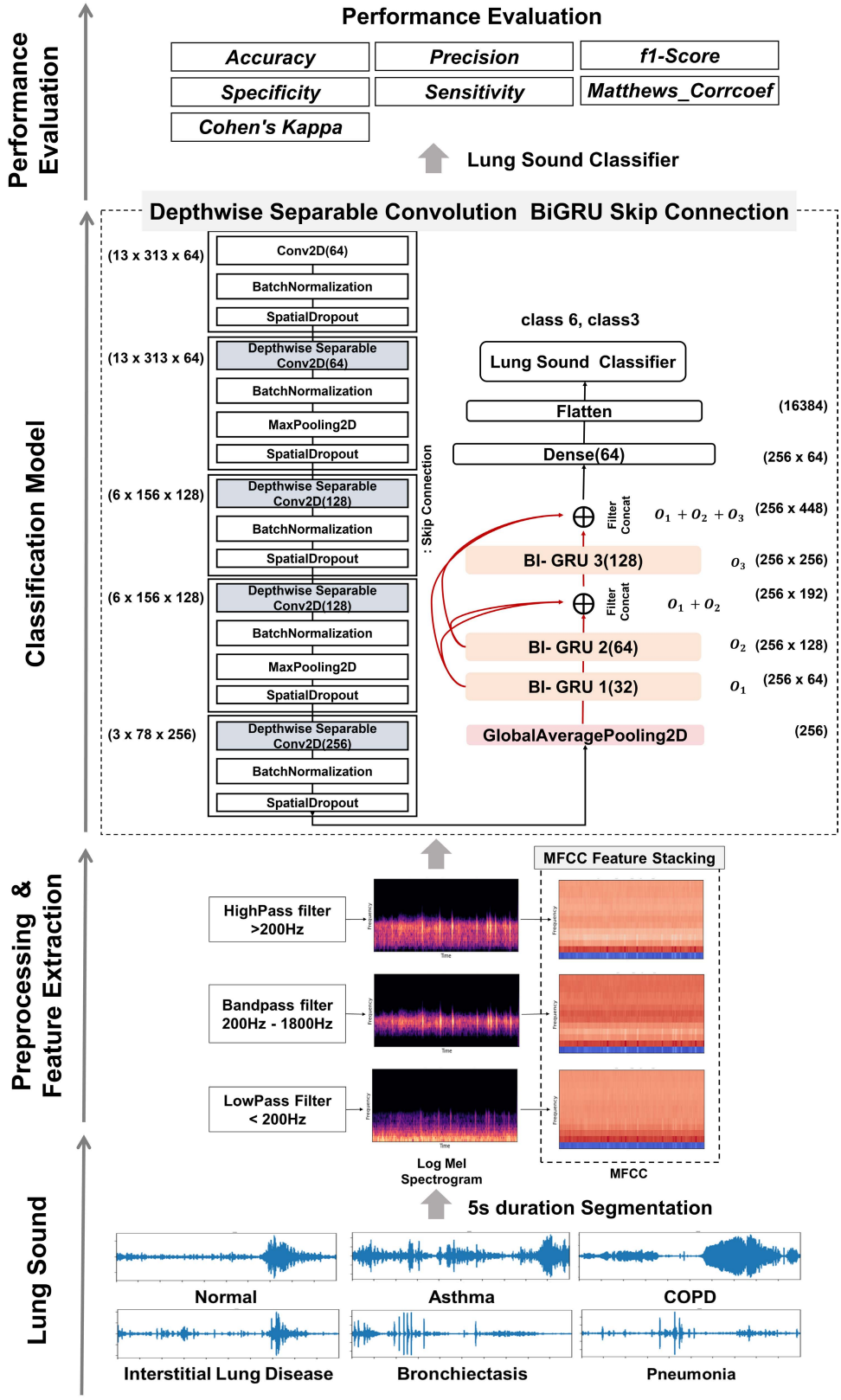

FIGURE 1. Architecture of the respiratory sound classification framework.

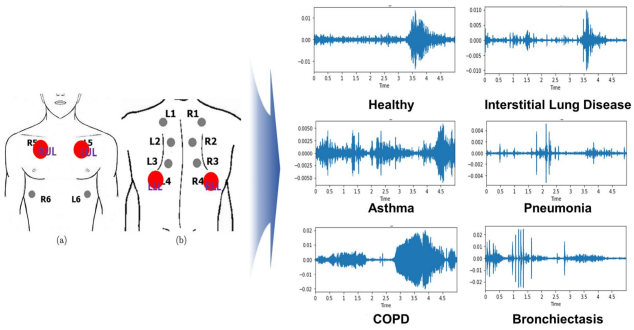

FIGURE 2. Patient measurement site and disease type: (a) anterior side and (b) posterior side [35].

III. METHOD

A. PROPOSED LEARNING FRAMEWORK OVERVIEW

Fig. 1 is the framework proposed in this study, and the procedure is as follows:

1) A 4 kHz sampling rate and 5 s segmentation were performed on the patient's 126 respiratory sounds.
2) Low pass (<200 Hz), band pass (200-1800 Hz), and high pass (>200 Hz) were used for normal and abnormal sounds to remove noise generated during measurement. After removing unnecessary noise, the breathing sound of the main band was obtained.
3) Respiratory sound feature extraction was performed using Log-Mel spectrogram-MFCC for three filters to extract breath sound information and feature stacking.
4) Disease and symptom classification for lung disease was performed using dense lightweight CNN-BiGRU skip connections. Abnormal respiratory symptoms are classified into 3 classes: normal, wheeze, and crackle. Accordingly, symptoms related to diseases were classified into 6 classes: normal, COPD, asthma, bronchiectasis (BRE), pneumonia, and interstitial lung disease (ILD).
5) The classification model was evaluated using seven performance indicators: accuracy, precision, sensitivity, specificity, f1-score, cohen's kappa, and matthews correlation coefficient.

B. DATASET

In this study, 126 normal and disease sounds were collected in cooperation with the allergy department in the participating hospital. Respiratory specialists collected the data from patients 19 years of age or older (inpatient and outpatient) who were being treated for lung disease by the Department of Allergy from November 2020 to May 2021 after passing the institutional review board (IRB) review in September 2020. As shown in Fig. 2, the specialist measured four anterior and posterior chest areas using a ''Littman 3200'' stethoscope [35]. The measured respiratory sound data is divided into normal, crackle, and wheeze [36]. Wheezing is characterized by a flute-like ''wheezing'' sound as air passes through a narrow airway.

TABLE 1. Disease information in the clinical dataset.

Name	Value
Number of recordings	126
Average recording duration	50 s / 1 min
Diagnosis	Normal (12) Wheeze: COPD (20), Asthma (23) Crackle: Bronchiectasis (25), Pneumonia (20), ILD (26)
Age	66±9
Height of participants	160.12±10.3
Weight of participants	60.5±11.4
BMI of participants	23.5±3.8
Basic Diseases	Hypertension, Diabetes, Cardiovascular disease, Arrhythmia, Chronic kidney failure, Cerebrovascular Cirrhosis of the liver, Chronic hepatitis, Blood cancer, Hyperlipidemia, Rhinitis, Solid tumors

TABLE 2. Clinical dataset.

Clinical Set		Train	Test	Total
Symptom	Disease			
Healthy	Normal	91	23	114
Wheeze	COPD	128	32	160
	Asthma	147	37	184
Crackle	Bronchiectasis	160	40	200
	Pneumonia	128	32	160
	Interstitial lung disease	163	40	203
Total		**817**	**204**	**1021**

COPD and asthma are representative diseases and appear when symptoms such as cough, sputum, and shortness of breath worsen. Crackle is characterized by a low-pitched or harsh sound [37] during inhalation. The representative diseases are pneumonia, bronchiectasis, and interstitial lung disease. BRE is characterized by sputum scraping. ILD is characterized by auscultation of crackling and dry sounds during inspiration. Information regarding the collected respiratory sounds is shown in Table 1.

The clinical dataset was set to 4000Hz because the sampling rate for each breath sound was different. The respiratory sound cycle was divided into 5 seconds based on inhalation and exhalation, according to the opinion of respiratory specialists. A total of 1021 data were obtained, and the data used for model training and evaluation are shown in Table 2.

C. PREPROCESSING AND FEATURE EXTRACTION

1) NOISE FILTERS: LOW, BAND, HIGH PASS

During auscultation of respiratory sounds, noise is generated from the friction of clothing and contact with the patient's

TABLE 3. Characteristics of respiratory sounds.

Symptom	Disease	Hz
Normal	Normal	150–1000 Hz
Wheeze	COPD, Asthma	>200 Hz
Crackle	Bronchiectasis, Pneumonia, Interstitial Lung Disease	200–2000 Hz

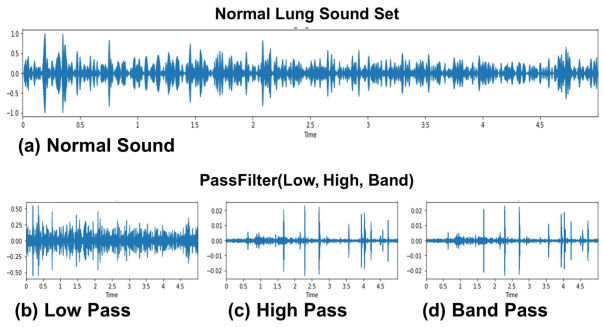

FIGURE 3. Preprocessing of respiratory sound: (a) normal sound, (b) low pass filter, (c) high pass filter, and (d) band pass filter.

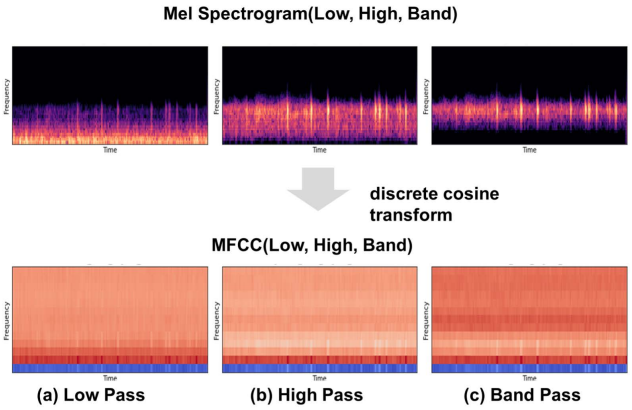

FIGURE 4. Respiratory sound feature extraction of the Mel spectrogram and the MFCC: (a) low pass filter, (b) high pass filter, and (c) band pass filter.

skin. Therefore, band, low, and high pass filters were applied to remove noise [38]. The band pass filter simultaneously removes low-frequency and high-frequency noise for a specific signal bandwidth, and the low pass filter extracts the low-frequency signal after removing the high-frequency signal's noise.

The high pass filter uses a high-frequency signal from which the low-frequency noise of a signal is removed [39]. As shown in Table 3, respiratory sounds are measured at 150–1000 Hz for normal sounds, over 200 Hz for wheezes, and 200–2000 Hz for crackles [40]. Therefore, this study set the pass filter for extracting the optimal breathing sound to 200Hz [41].

Low pass filter was applied for sounds of 200 Hz or less, high pass filter was applied for 200 Hz or more sounds, and band pass filter was applied for 200–1800 Hz. The results of noise removal for each disease are shown in Fig. 3. In a previous study, one pass filter was applied to remove noise [31], but in this study, all three filters were applied, as shown in Fig. 4, to implement feature stacking.

TABLE 4. Characteristics of the Mel spectrogram and the MFCC.

Mel Spectrogram	Value	MFCC	Value
Number of Mel Bins	64	Number of MFCCs	13
FFT Window Size	256	FFT Window Size	256
Hop Length	64	Hop Length	64

2) LOG-MEL SPECTROGRAM-MFCC

To extract the characteristics of respiratory sounds, we used the Mel spectrogram and MFCC, which are widely used in the field of speech signals. A spectrogram is a feature extraction method that divides an audio signal into specific sections and analyzes the spectrum in each section. It visually expresses the voice signal by ascertaining the frequency intensity according to the change in time [42]. A Mel spectrogram includes information that is obtained after converting the spectrogram's physical frequency to the Mel-scale by reflecting the characteristics of the human hearing organ. The MFCC performs discrete cosine transform (DCT) on the Mel spectrogram [43]. This process reduces abnormal values in the respiratory sounds and removes noise by compressing it [44]. Therefore, in this study, the MFCC was applied, and the MFCC attribute extraction estimates had several stages, which are explained below:

1) We calculated the spectrogram after applying STFT to the respiratory sound.
2) After applying the Mel filter bank from (1) to the spectrogram, we obtained the Mel spectrogram. The Mel scale frequency(F_{mel}) formula is as follows:

$$F_{mel} = 2595 log_{10}(1 + \frac{f}{700}) \tag{1}$$

3) After normalizing the Mel spectrogram, we used (2) to scale the signal to the filter bank's value and convert the dB unit. This process is defined by the Log-Mel spectrogram [45]. Fig. 5 is the Mel spectrogram of the disease class. The formula is as follows:

$$10 log_{10}(\frac{s}{ref}) \tag{2}$$

4) In our experiment, the MFCC was extracted using DCT transformation. The shape and size of the MFCC were composed of 313 frames with a window length of 64 ms ($0.064 \times 4{,}000 = 256$) overlapped by 75% in consideration of the respiratory sound cycle. The extracted results are shown in Fig. 4. As for the hyperparameters, which were efficiently applied to extract the respiratory sound characteristics in Table 4.
5) Feature stacking was performed on the extracted MFCCs, as shown in Fig. 6.

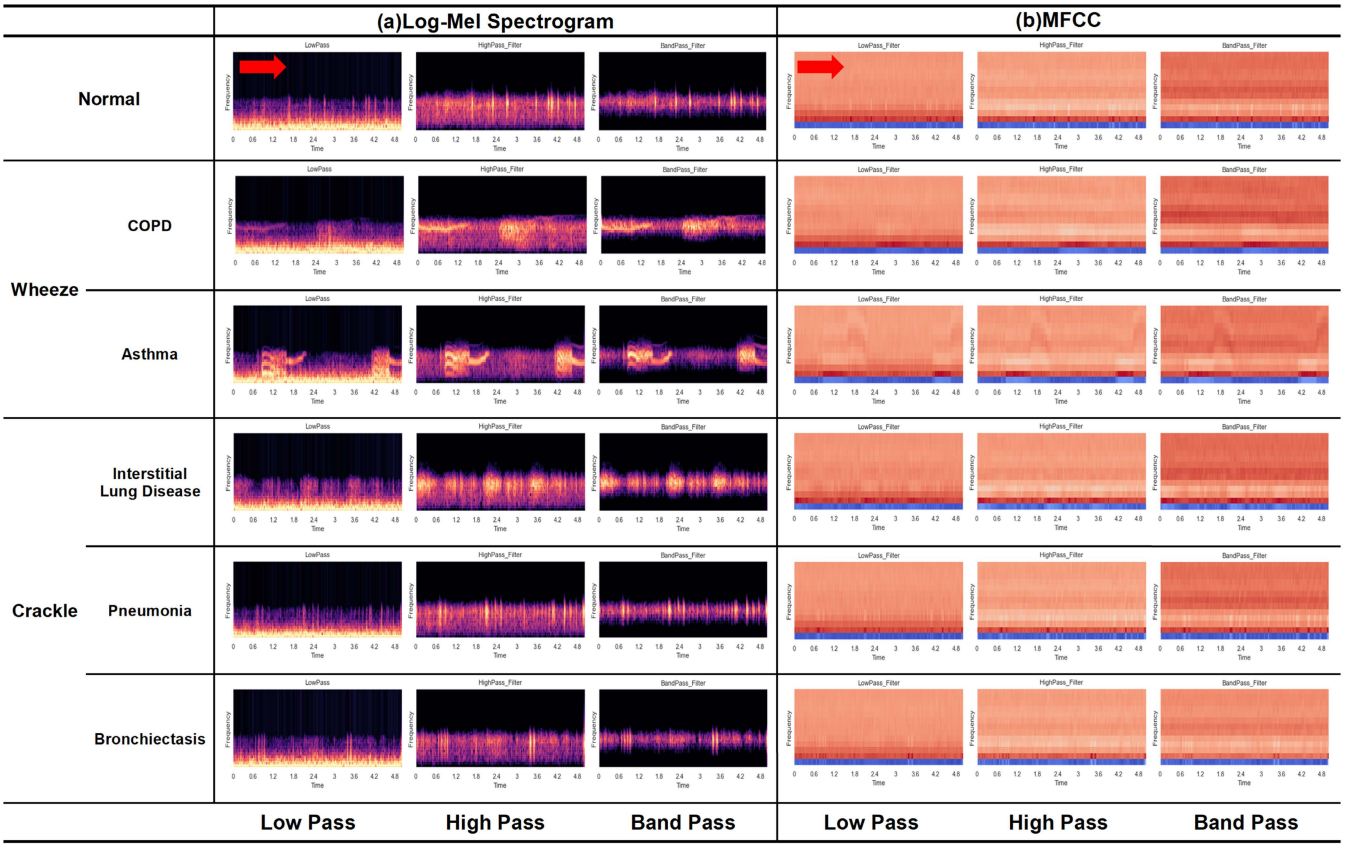

FIGURE 5. Respiratory sound(5s) feature extraction of the (a) Log-Mel spectrogram and (b) MFCC with Normal, Wheeze, and Crackle. Normal is a general respiratory sound. Wheeze is characterized by musical, high-pitched sound that is heard on exhalation and can be heard in patients with COPD or asthma. Crackle is characterized by a nonmusical, short, and explosive sound which can be heard in patients with interstitial lung disease, pneumonia, or bronchiectasis.

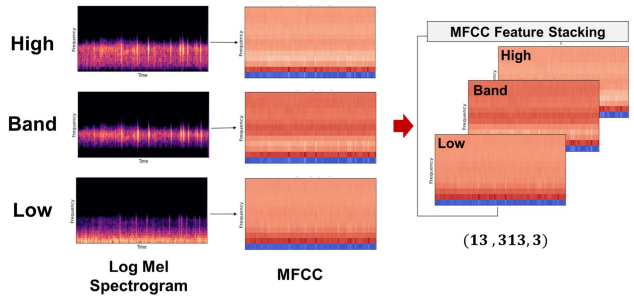

FIGURE 6. MFCC with feature stacking. The size of the MFCC was converted to 13 × 313 and used as an input to a CNN with 3 channels.

D. DENSE DEPTHWISE SEPARABLE CNN-BiGRU SKIP CONNECTIONS

1) DEPTHWISE SEPARABLE CONVOLUTION

A CNN is a convolutional deep learning method composed of a convolution layer and a pooling layer. It has been applied to speech and image processing, and it plays an important role in feature extraction [46], [47]. Therefore, in this study, depthwise separable convolution was applied for efficient calculation and weight reduction of the respiratory sound classification model.

This convolution structure is used in MobileNetv1 [48] as a method that reduces the CNN convolution cost. The model is depicted in Fig. 8, and its features are composed of a depthwise structure that separates the feature map for each channel and applies a 1 × 1 pointwise convolution

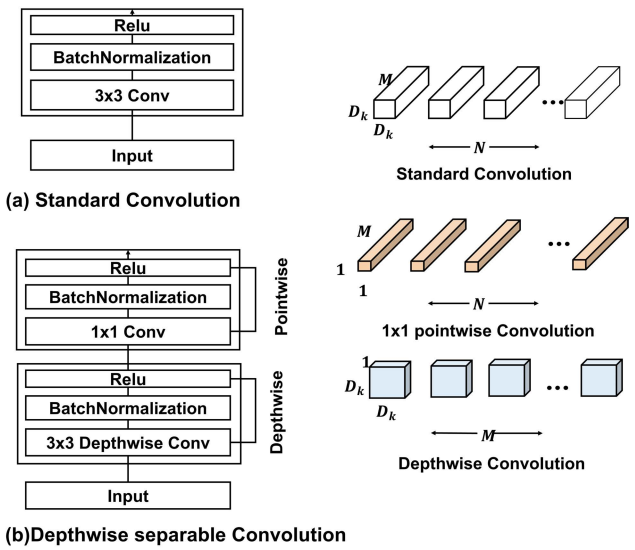

FIGURE 7. (a) Standard convolution and (b) depthwise separable convolution, the kernel size D_k, the number of output channels N, and the number of input channels M [48].

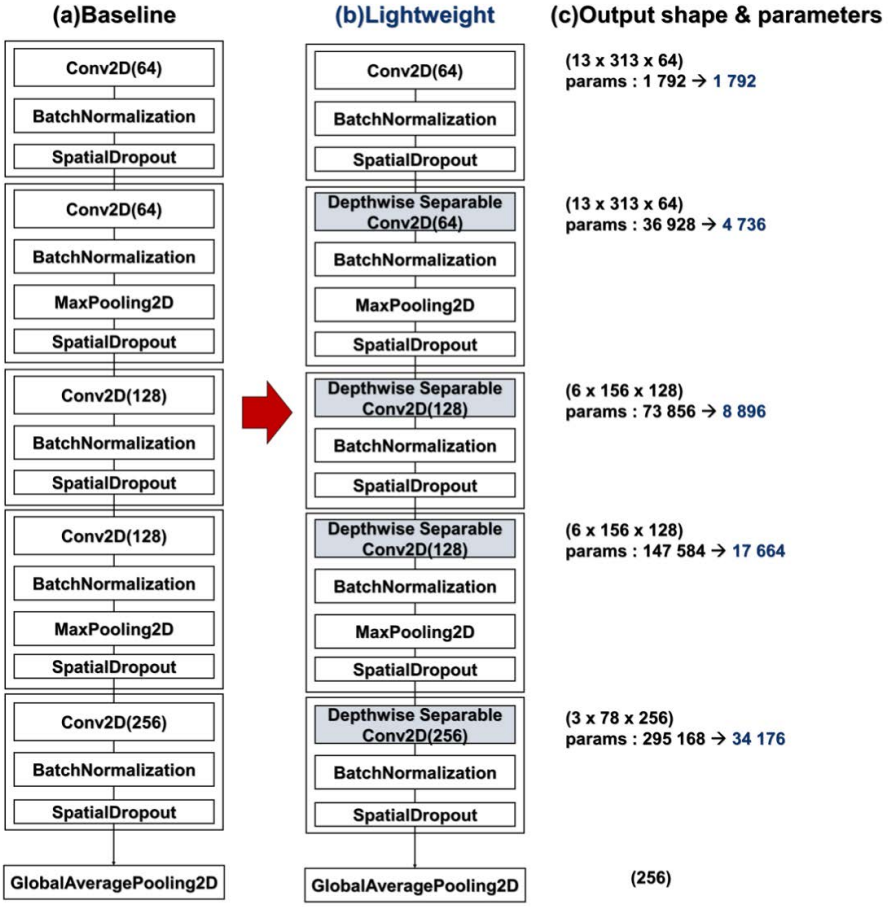

FIGURE 8. Architecture of our model according to (a) the baseline, (b) the lightweight version, and (c) output shape and parameters.

that merges multiple channels into one new channel [48]. This process was used in the mobile environment, and the structure of the applied model is shown in Fig. 7. The model reduced weights by changing conv2D to a depthwise separable convolution.

2) GRU

The GRU is an RNN-based method proposed by Cho [49]. It is useful for processing time series data as it remembers past information based on sequential data and reflects current information. This method performs a role similar to LSTM [50], but it solves the LSTM long-term dependency problem by efficiently processing the weights to be learned with a simpler structure that employs an update gate and a reset gate that combines the LSTM input and forget gate. The GRU flow chart is shown in Fig. 9.

a: RESET GATE (r_t)

Multiply both weights (W_r),(U_r) of the previous time point (h_{t-1}) and the current time point (x_t), and input the sum of the two operations into the sigmoid function to change (3) to output a value between 0 and 1. This function partially deletes past information and determines the degree to which the hidden state (h_{t-1}) of the previous time is reflected. The r_t formula is as follows:

$$r_t = \sigma(w_r x_t + U_r h_{t-1} + b_r) \tag{3}$$

b: UPDATE GATE (z_t)

When the current information (x_t) is input, it is multiplied by the weight of the viewpoint (W_z), and the previous time (h_{t-1}) is multiplied by the weight of the viewpoint (U_z). Then, the current and previous information's combined value is input into the sigmoid function and output as a value between 0 and 1 using (4) to determine how long the previous information will be maintained. The z_t formula is as follows:

$$z_t = \sigma(w_z x_t + U_z h_{t-1} + b_z) \tag{4}$$

c: CANDIDATE ($\tilde{h}_t$)

The candidate performs a multiplication operation on the result of the reset gate in a hidden state. In (5), h_{t-1} denotes $t-1$ of the hidden layer state. The $\tilde{h}_t$ formula is as follows:

$$\tilde{h}_t = tanh(w_h x_t + U_h(r_t * h_{t-1} + b_h) \tag{5}$$

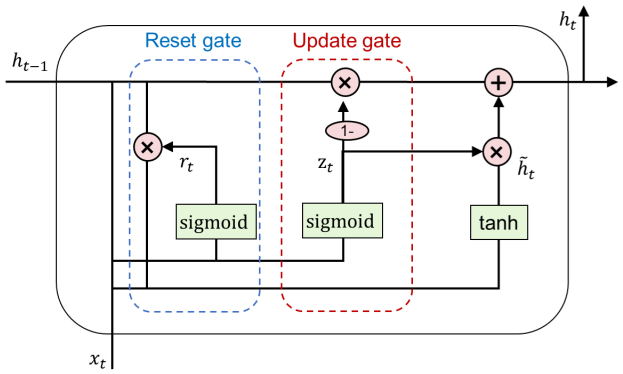

FIGURE 9. GRU architecture.

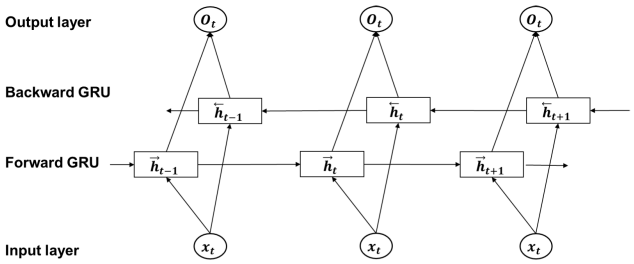

FIGURE 10. BiGRU diagram [17].

d: HIDDEN STATE (h_t)

Combining the update gate of (6) and the candidate ($\tilde{h}_t$), the hidden state of the current time is calculated. The h_t formula is as follows:

$$h_t = (1 - z_t) * h_{t-1} + zt * \tilde{h}_t \tag{6}$$

3) BIGRU

In this study, the BiGRU was used to determine the relationship between the information on previous and current respiratory sounds based on the GRU by considering the temporal period characteristics. This strategy has an advantage in terms of ascertaining the characteristics of the respiration sounds, as there is a close relationship between the previous and late states of respiration. In contrast, a GRU only remembers information at a specific point in time [7], [17]. The BiGRU is a neural network composed of forwarding and reversing directions, and it is connected to the output layer, as shown in Fig. 10.

A BiGRU (h_t) is composed of a forward GRU ($\overrightarrow{h}_t$) and reverse GRU ($\overleftarrow{h}_t$). When the input (x_t) is used as an input value, the forward direction ($\overrightarrow{h}_t$) of (7) and reverse direction ($\overleftarrow{h}_t$) of (8) are output. w_t and v_t are calculated similarly to h_t in (9) by summing the weights for each direction in both directions of $\overrightarrow{h}_{t-1}$and $\overleftarrow{h}_{t-1}$. b_t is the bias of the hidden layer's state at t.

$$\overrightarrow{h}_t = GRU(x_t, \overrightarrow{h}_{t-1}) \tag{7}$$
$$\overleftarrow{h}_t = GRU(x_t, \overleftarrow{h}_{t-1}) \tag{8}$$
$$h_t = (w_t \overrightarrow{h}_t + v_t \overleftarrow{h}_t + b_t) \tag{9}$$

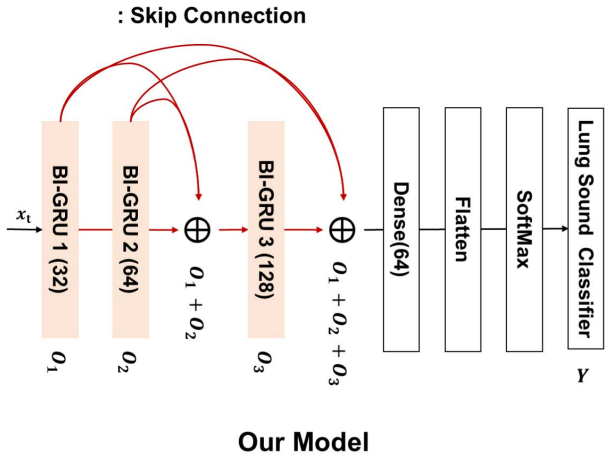

FIGURE 11. Architecture of skip connections with BiGRU.

4) DENSE BIGRU SKIP CONNECTIONS NETWORK

Dense BiGRU skip connections were applied to prevent model overfitting. When learning a deep learning model, as the stacked layers deepen, the problem of gradient vanishing occurs, and as a result, learning does not occur. This phenomenon leads to overfitting because, in many neurons, the gradients become too small or too large when reaching the lower layers during backpropagation [51]. Therefore, for this purpose, skip connections are applied to prevent overfitting. All the features from the previous layer relate to the features of the current layer to deliver continuous information, thereby preventing the loss of feature information in the respiratory sounds. The formula for skip connections is that the x_t is input to the BiGRU of (10), and the output (O_t) is calculated by (10)–(13). Fig. 11 depicts our process. In addition, the experimental results are compared with BiLSTM in Section V.

$$O_1 = GRU_1(x_t) \tag{10}$$
$$O_2 = GRU_2(O_1) \tag{11}$$
$$O_3 = GRU_3(O_1 + O_2) \tag{12}$$
$$Y = softmax(flatten(Dense(O_1 + O_2 + O_3))) \tag{13}$$

IV. EXPERIMENT

5) HYPERPARAMETER OF EXPERIMENTS SETUP

The experimental environment consisted of a GeForce 1080TI (NVIDIA), AMD Ryzen 7 2700X 8-core processor (CPU), and 64 GB RAM. Python 3.8 and TensorFlow 2.4.0 were also used for the experiment. The hyperparameters of the learning model were a batch size of 8, the number of epochs was 1000, and the learning rate was 3e-4. These values are shown in Table 5. The loss function used cross-entropy, and the optimization function used the Adam optimizer. Cross-validation was performed five times by dividing the dataset for training and test validation of the model into 80% training and 20% test samples [52], [43]. This setup can avoid the problem of overfitting bias to the evaluation set when training the model.

TABLE 5. Hyperparameter of experiments.

Parameter Names	Values
Spatial Dropout Rate	0.05 / 0.1
Learning Rate	3e-04
Optimizer	Adam
Batch Size	8
Number of Epochs	1000
Cross-Validation	5

Our CNN uses 64, 64, 128, 128, and 256 depthwise separable convolutions, pooling layers (maxpooling and global average pooling), batch normalization, and spatial dropout. Dropout [53] transforms the node value of the hidden layer into a random probability of 0 when learning a deep learning network. However, there is a correlation problem between channels in some spatial features extracted in the convolution layer. Spatial dropout was applied, and the hidden layer sizes of the BiGRU were set to 32, 64, and 128.

6) MODEL EVALUATION

The performance of the model has seven evaluators: accuracy (14), sensitivity (15), specificity (16), precision (17), the f1-score (18), cohen's kappa (19), and the matthews correlation coefficient (MCC) (20). The model was assessed using the evaluation index. The confusion matrix of the actual and predicted values was evaluated as true positive (TP), true negative (TN), false positive (FP), or false negative (FN) [54].

Accuracy is a measure of whether a model provides correct classifications and is most frequently used in performance evaluation. Specificity(recall) is the probability of predicting a disease if there is an actual disease and predicting the absence of disease if there is no disease. Performance indicators for sensitivity and specificity are used when developing diagnostic kits for influenza and other diseases, and the higher the performance, the higher the reliability. Precision is the probability of judging an actual disease when a disease is predicted, and the f1-score is the harmonized average probability of precision and recall [55].

Cohen's kappa, proposed by Cohen in 1960, measures the degree of agreement between two observers. In (19), p_o is the probability of agreement among the raters, and p_e is the probability of receiving a consistent evaluation from the rater [56]. The MCC is an index used for the classification evaluation of multiple classes, and the accuracy is between -1 and 1. The closer the accuracy is to 1, the more similar the observation results are judged to be [57].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

$$Sensitivity = Recall = \frac{TP}{TP + FN} \tag{15}$$

$$Specificity = \frac{TN}{FP + TN} \tag{16}$$

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{18}$$

$$cohen'sKappa(K_C) = \frac{p_o - p_e}{1 - P_e} \tag{19}$$

$$MCC(MatthewsCorrelationCoefficient) = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{20}$$

V. RESULTS

This section compared the model's performance for class 6 (disease) and class 3 (symptom).

TABLE 6. Results for class 6 using the standard CNN-BiGRU (1 stack), the baseline model (with skip connections) and the lightweight model (skip connections and depthwise separable convolution).

Class 6(%)	Standard	Ours	
	CNN-BiGRU	Baseline	Lightweight
Accuracy	91(±3.2)	**92.6(±1.3)**	92.3(±2.5)
Precision	91.2(±3.2)	**92.8(±1.4)**	92.0(±2.6)
Sensitivity	90.6(±3.2)	**92.3(±1.3)**	92.1(±2.6)
Specificity	98.2(±0.6)	**98.5(±0.3)**	98.5(±0.5)
F1-Score	90.7(±3.3)	**92.4(±1.2)**	91.9(±2.6)
Cohen's Kappa	89.1(±3.9)	**91.0(±1.5)**	90.7(±3.0)
MCC	89.2(±3.8)	**91.1(±1.6)**	90.7(±3.0)

A. RESULTS FOR CLASS 6 (DISEASE)

Table 6 shows the results of 5-fold cross-validation on the methodology proposed for class 6 (disease). Our proposed baseline model applied to skip connections, and high performance was confirmed with an average accuracy of 92.6%, precision of 92.8%, sensitivity of 92.3%, specificity of 98.5%, F1-score of 92.4%, Cohen's kappa of 91%, and MCC of 91.1%. The lightweight model employs depthwise separable convolution, and it achieved an accuracy of 92.3%, which is 0.3% lower than the baseline performance. However, it still demonstrated high performance. When comparing the standard deviation of the mean value of 5 cross-validation, the baseline exhibited the smallest standard deviation value, 1.3%, compared to other models. In addition, by maintaining a stable performance, the effect of the model proposed in this study was confirmed.

Our model's performance is superior to the standard (CNN-BiGRU) because the GRU layer was added to improve the learning performance. The model's performance was maintained by stabilizing the model while updating the GRU network by solving the vanishing gradient through skip connections to the stack layer; our model also transfers information about the previous and current respiratory sounds. This strategy prevents overfitting and increases the accuracy by 1.6%. In addition, the precision, sensitivity, specificity, and f1-score of our model's results for the baseline in Table 6 are presented in Table 7 and Table 8.

TABLE 7. Results for the baseline model with precision, sensitivity, specificity, and f1-score.

Class 6(%)	Precision	Sensitivity	Specificity	F1-score
Bronchiectasis(BRE)	86.7	93.0	96.3	89.5
COPD	95.4	88.1	99.2	91.5
Interstitial lung disease(ILD)	95.8	97.1	98.9	96.4
Normal(NL)	88.7	91.2	98.5	89.7
Pneumonia	92.6	91.9	98.6	92.2
Asthma	97.7	92.4	99.5	94.9
average	**92.8**	**92.3**	**98.5**	**92.4**

TABLE 8. Results for the lightweight model with precision, sensitivity, specificity, and f1-score.

Class 6(%)	Precision	Sensitivity	Specificity	F1-score
Bronchiectasis(BRE)	92.7	92.5	98.2	92.5
COPD	91.9	90.6	98.5	91.2
Interstitial lung disease(ILD)	93.7	94.6	98.4	94.1
Normal(NL)	85.3	90.4	98.0	87.7
Pneumonia	91.4	91.9	98.4	91.5
Asthma	96.6	92.4	99.3	94.4
average	**92.0**	**92.1**	**98.5**	**91.9**

When comparing the results of each disease class, our baseline demonstrated good predictive performance for each disease when performance indicators of precision, sensitivity, specificity, and f1-score were compared.

In terms of the f1-score, for the baseline model, the performance for pneumonia was 92.2%, and COPD was 91.5%; for the lightweight model, the performance for pneumonia was 91.5%, and COPD was high at 91.2%. Pneumonia is a disease with a high death rate among patients with respiratory diseases, and early diagnosis by the proposed model will make it possible to prepare for the prediction of COVID-19 symptoms similar to infectious diseases and acute respiratory infection (pneumonia) [58]. Fig. 12 is the confusion matrix for the best result from each model. When the results of the three models were compared, even for the best model, the baseline's results were classified without being biased by a specific class, and the model's superiority could be confirmed.

B. RESULTS FOR CLASS 3 (SYMPTOM)

In the case of class 3 (symptom), in Table 2, the respiratory sounds from 6 disease groups were grouped into three symptom groups, and the experimental results were confirmed. Table 10 shows the average of 5-fold cross-validation for class 3. For this class, the baseline model demonstrated excellent performance with an accuracy of 94.5%, precision of 93.2%, sensitivity of 92.7%, specificity of 96.7%, f1-score of 92.8%, Cohen's kappa of 90.4%, and MCC of 90.5%. The accuracy of the lightweight model was 94.6%, which increased by 0.1% compared to the baseline performance, demonstrating its capabilities. However, the sensitivity and

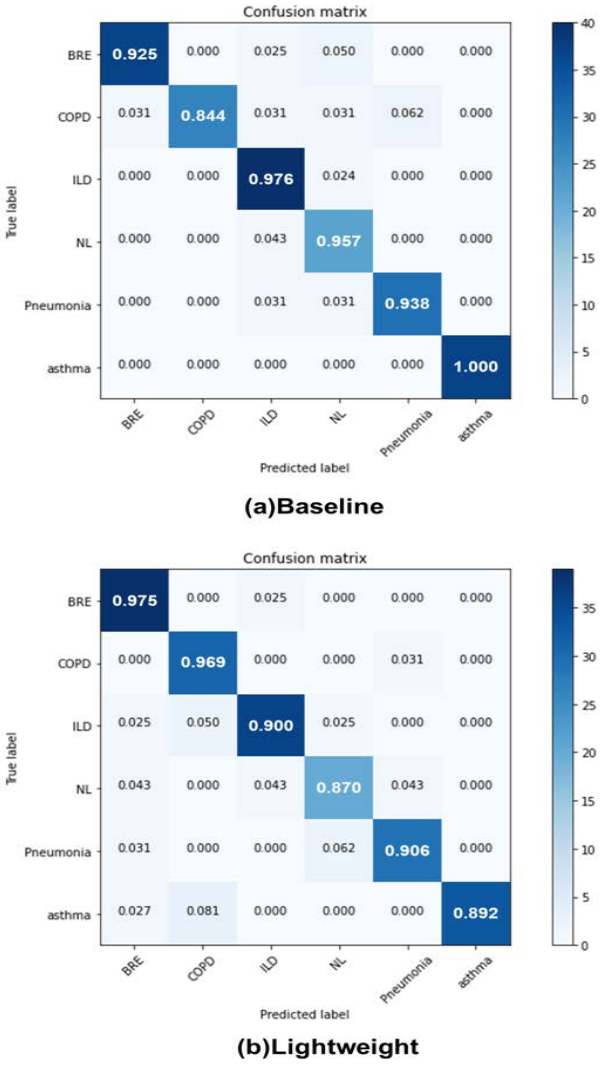

FIGURE 12. The confusion matrix for (a) the baseline model and (b) the lightweight model (class 6).

TABLE 9. Results for class 3 using the standard CNN-BiGRU (1 stack), the baseline model (with skip connections) and the lightweight model (skip connections and depthwise separable convolution).

Class 3(%)	Standard	Ours	
	CNN-BiGRU	Baseline	**Lightweight**
Accuracy	93.6(±2.6)	94.5(±2.7)	**94.6(±1.7)**
Precision	91.6(±4.1)	93.2(±4.8)	**93.3(±2.4)**
Sensitivity	92.1(±3.5)	**92.7(±3.8)**	91.8(±1.1)
Specificity	96.4(±1.6)	96.7(±1.5)	**96.8(±0.9)**
F1-Score	91.6(±3.5)	**92.8(±3.9)**	92.5(±1.6)
Cohen's Kappa	88.9(±4.5)	90.4(±4.8)	**90.5(±2.9)**
MCC	89.0(±4.4)	90.5(±4.9)	**90.6(±2.9)**

f1-score, the baseline showed higher performance, confirming the superiority of the baseline classification model overall.

Table 10 and Table 11 summarize the performance evaluation regarding the presence or absence of abnormalities

TABLE 10. Results for the baseline model with precision, sensitivity, specificity, and f1-score.

Class 3(%)	Precision	Sensitivity	Specificity	F1-score
Normal	87.6	88.6	98.2	87.7
Crackle	94.5	96.8	93.0	95.6
Wheeze	97.6	92.7	98.8	95.1
average	**93.2**	**92.7**	**96.7**	**92.8**

TABLE 11. Results for the lightweight model with precision, sensitivity, specificity, and f1-score.

Class 3(%)	Precision	Sensitivity	Specificity	F1-score
Normal	89.2	83.3	98.7	86.0
Crackle	95.1	95.9	93.9	95.5
Wheeze	95.7	96.2	97.8	95.9
average	**93.3**	**91.8**	**96.8**	**92.5**

in respiratory symptoms, not diseases as in class 6 (disease). Table 10, Table 11, and Fig. 13 summarize the performance evaluation regarding the presence or absence of respiratory abnormalities, not diseases as in class 6 (disease). Each class showed high predictive performance for symptom classification, indicating significant diseases such as crackles and wheezes. Therefore, the model proposed in this study can judge respiratory symptoms for both medical staff and patients. The model's superiority was confirmed through the results for both class 6 and class 3.

C. COMPARATIVE EXPERIMENTS

Comparative experiments were conducted based on class 6 (disease).

1) COMPARISON OF THE BASELINE AND LIGHTWEIGHT MODELS

Table 12 summarizes the baseline and lightweight model indicators and suggests a lightweight model that can be installed in a mobile environment. The indices of weight reduction were compared with the number of parameters, size of the model, computation time, inference time, and multiply-adds (MAdds) according to the input size. MAdds is calculated using standard convolution in (21) and depthwise separable convolution in (22). The computational parameter $D_k \times D_k$ is the size of the kernel, N is the number of filters, M is the number of channels of the input data, and $D_F \times D_F$ is the size of the input data [48]. MAdds calculates only the convolution cost.

$$D_k * D_k * N * M * D_F * D_F \tag{21}$$

$$D_k * D_k * M * D_F * D_F + (M * N * D_F * D_F) \tag{22}$$

As shown in Fig. 14, the parameters decreased by more than 50%, from 988 870 to 500 806. The model's size decreased by more than 49%, from 11 791 to 6 048, in the lightweight model compared to the baseline. Additionally, when the input size was changed and compared, the number of parameters was the same, but the accuracy of both MAdds and the model tended to decrease. Therefore, the input size of 13 × 313 proposed in this study demonstrated the highest model performance with an accuracy of 92.3%. Compared with the baseline, the effect of weight reduction was confirmed, and the performance was maintained.

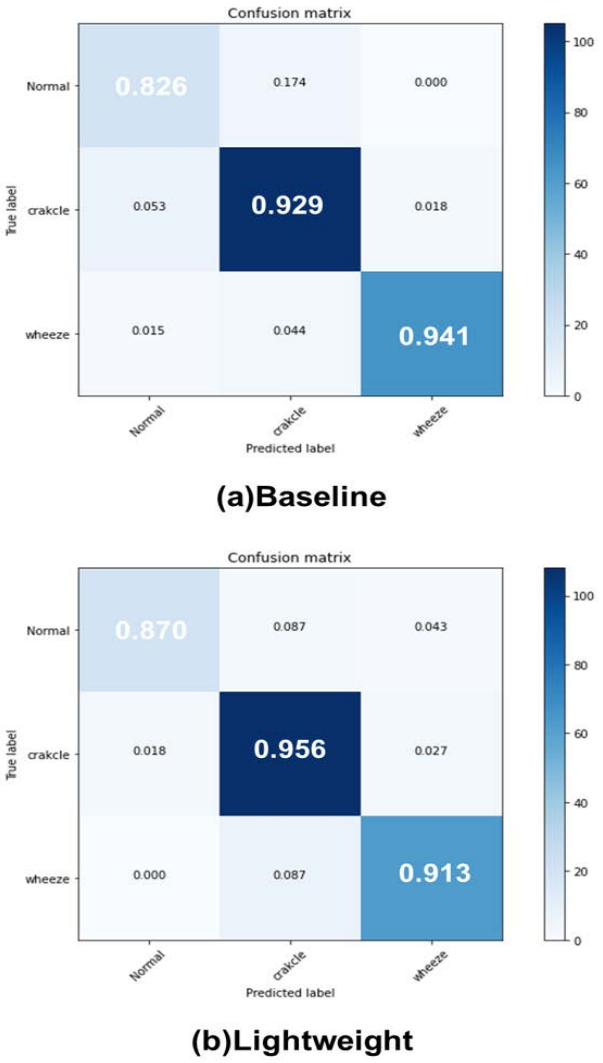

FIGURE 13. The confusion matrix for (a) the baseline model and (b) the lightweight model (class 3).

TABLE 12. Comparison of the baseline and lightweight models for class 6 (disease).

Input Size	Baseline	Lightweight		
	13 × 313	13 × 313	13 × 157	13 × 79
Number of Parameters	988,870	500,806		
Size of Model(KB)	11,791	6,084		
Computation Time(MS)	46	50		
Inference time(MS)	202	195	185	180
MAdds(M)	109	64	32	16
Accuracy(%)	92.6 (±1.3)	92.3 (±2.5)	90.9 (±1.7)	84.8 (±2.0)

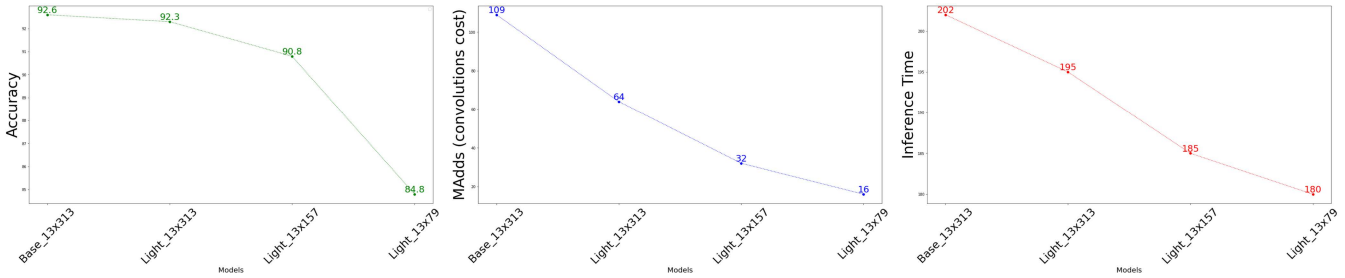

FIGURE 14. Comparison of (a) accuracy, (b) MAdds, and (c) inference time.(Base 13 × 313, Light 13 × 313, Light 13 × 517, and Light 13 × 79).

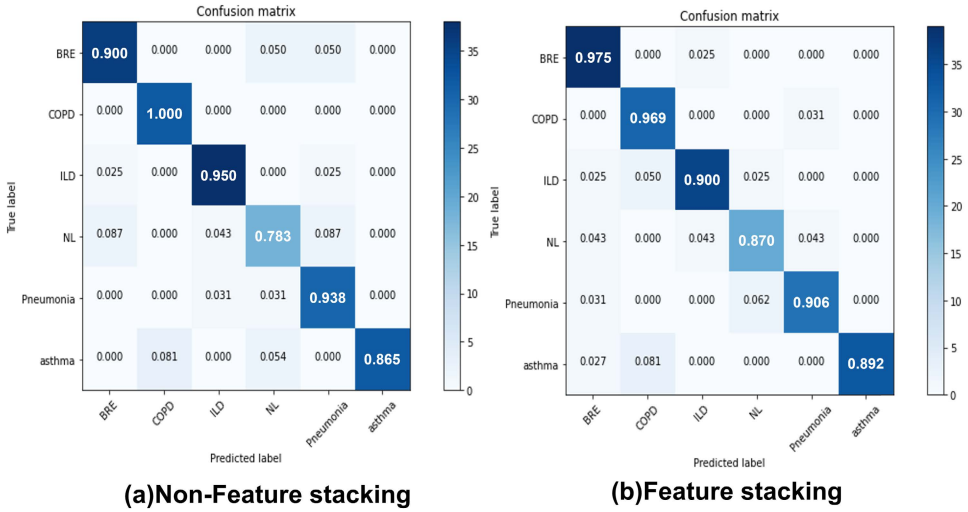

FIGURE 15. The confusion matrix for (a) non-feature stacking and (b) feature stacking.

TABLE 13. Comparison of feature stacking with lightweight model(%).

	Non-Feature Stacking (Band pass Filter)	Feature Stacking
Accuracy	90.6(±1.7)	**92.3(±2.5)**
Precision	90.6(±1.7)	**92.0(±2.6)**
Sensitivity	90.2(±1.7)	**92.1(±2.6)**
Specificity	98.1(±3.0)	**98.5(±0.5)**
F1-Score	90.3(±1.7)	**91.9(±2.6)**
Cohen's Kappa	88.6(±2.0)	**90.7(±3.0)**
MCC	88.7(±2.0)	**90.7(±3.0)**

2) COMPARISON OF FEATURE STACKING

We compared our method with the experimental results of non-feature stacking to confirm the effect of feature stacking proposed in this study. The results are shown in Table 13. In the non-feature stacking, only the band pass filter was applied from among the three filters. Our method, which applies stacking, improved the performance by 2%, with an accuracy of 92.2% and an f1-score of 91.9%. In addition, feature stacking achieved a greater than 90% value in all performance. The effect of the stacking feature for each of the three filters, band pass, low pass, and high pass, was confirmed.

This result occurs because stacking emphasizes the required respiration information and diversifies the bandwidth. The results for the confusion matrix shown in Fig. 15 show that the model tends to be somewhat confused when classifying lung diseases in the case of non-feature stacking. Although BRE and pneumonia were confused when classifying similar crackle symptoms, our model provided correct classifications.

3) COMPARISON OF BIGRU AND BILSTM USING A CNN

A GRU is a model in which the LSTM structure is improved using a reset gate and an update gate. It has the advantage of similar or better performance in voice and signal modeling [59]. In Table 14, when comparing the model's performance with one BiGRU and BiLSTM layer added to standard CNN, it was confirmed that the GRU had similar or improved model performance and a similar or fewer number of parameters compared to LSTM. Our models' skip connections and experiment were compared by changing the number of hidden layers of BiLSTM and the BiGRU according to standard convolution and depthwise separable convolution. Although our lightweight model applies to skip connections by adding a hidden layer, our baseline model

TABLE 14. Comparison of bigru and bilstm using the standard and lightweight cnn(%).

	Baseline (Standard CNN)				Lightweight (Depthwise Separable Convolution)			
	BiLSTM(1 stack)	BiGRU(1 stack)	BiLSTM(skip)	Ours (skip)	BiLSTM(1 stack)	BiGRU(1 stack)	BiLSTM(skip)	Ours(skip)
Accuracy	90.5	91.5	92.4	**92.6**	91.6	91.1	92.3	**92.3**
Precision	90.4	91.2	92.2	**92.8**	91.4	91.1	92.0	**92.0**
Sensitivity	90.2	90.6	92.1	**92.3**	91.5	90.8	92.1	**92.1**
Specificity	98.1	98.2	98.5	**98.5**	98.3	98.2	98.5	**98.5**
F1-Score	90.2	90.7	92.1	**92.4**	91.3	90.8	91.9	**91.9**
Cohen's Kappa	88.5	89.1	90.8	**91.0**	89.8	89.2	90.7	**90.7**
MCC	88.6	89.2	90.8	**91.1**	89.9	89.3	90.7	**90.7**
Params	669,062	667,078	1,088,390	**988,870**	180,998	179,014	600,326	**500,806**

TABLE 15. Clinical set using ICBHI.

Clinical Set using ICBHI		Train	Test	Total
Symptom	Disease			
Healthy	Normal	231	23	254
Wheeze	COPD	553	32	585
	Asthma	151	37	188
Crackle	Bronchiectasis	224	40	264
	Pneumonia	276	32	308
	Interstitial lung disease	163	40	203
Total		**1598**	**204**	**1802**

uses 500 806 fewer parameters than the baseline BiLSTM (1 stack with an accuracy of 90.5%) and BiGRU (1 stack with an accuracy of 91%), and it achieves an accuracy of 92.3% while being lighter weight. This experiment confirms that the application effect of depthwise separable convolution and a GRU is simpler than LSTM.

4) COMPARISON USING THE TRAIN ICBHI DATASET

In addition to the clinic data used to verify the stability and reliability of the lung disease classification model proposed in this paper, an additional experiment was performed by mixing the ICBHI dataset used in previous studies. We used data collected from "LittC2SE" and "Litt3200" among the ICBHI datasets, and the composition of the dataset used for the experiment is shown in Table 15 and Fig. 16.

Table 16 shows the results of comparing the performance of the proposed model. Similar results were obtained within a range of ±1–2% in terms of key performance evaluation indicators, including the model's accuracy when learning with a mixed model and learning only from clinic data and ICBHI. According to these results, the model proposed in this study was sufficiently trained using only clinic data, demonstrating its performance stability.

VI. DISCUSSION

In this study, we proposed a lightweight lung disease classification model due to removing noise and improved feature stacking using respiratory sounds directly labeled by a respiratory specialist. As a characteristic of the study, three filters were stacked to diversify the band information and concentration of respiration information, and temporal and spatial information related to the respiration sounds were efficiently learned by combining the CNN and BiGRU.

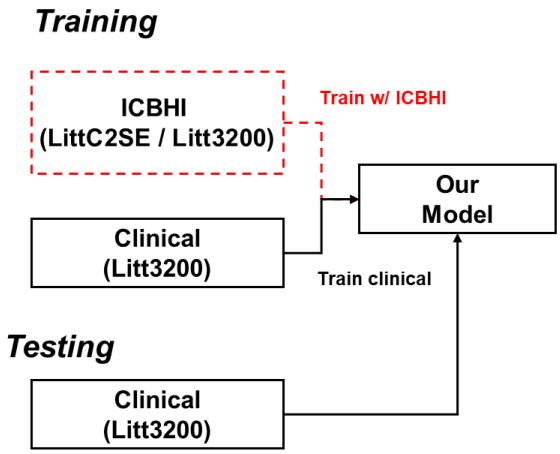

FIGURE 16. Training with ICBHI.

TABLE 16. Results of training with ICBHI dataset(%).

	Training w/ICBHI	Training Clinical
Accuracy	90.7	91.7
Precision	90.6	91.3
Sensitivity	90.3	91.2
Specificity	98.1	98.4
F1-Score	90.4	91.1
Cohen's Kappa	88.8	89.9
MCC	88.8	90.0

The performance of the proposed model obtained an accuracy of 92.3%, sensitivity of 92.1%, and specificity of 98.5% for the disease. It showed more than 90% performance for pneumonia, similar to COVID-19. In the case of symptoms, the model's performance confirmed superiority with an accuracy of 94.6%, sensitivity of 91.8%, and specificity of 96.8%. Comparative experiments on weight reduction (input size, convolution), feature stacking, GRU(LSTM) skip connections, and comparative learning experiments on public data provided by ICBHI have proven robustness even in datasets directly measured by specialists.

VII. CONCLUSION

To classify lung diseases, proposing a lightweight-based model using depthwise separable convolution is expected to suggest the possibility of a lightweight device and help in the

early diagnosis of disease classification. However, the data were limited in collecting many respiratory sounds due to COVID-19. When it is difficult to contact patients remotely, our model can build more accurate labeling and artificial intelligence modeling than public datasets in a pandemic. This study is expected to be an artificial intelligence solution that can solve the increased workload for medical staff due to the spread of infectious respiratory diseases and provide home care providers with the ability to check for disease in real-time. In the future, we plan to conduct XAI research that analyzes the relationship between disease and underlying disease by acquiring additional respiratory sounds and biosignal data from COVID-19 patients to identify the cause of the disease.

REFERENCES

[1] Global Asthma Network. (2014). *The Global Asthma Report 2014*. Auckland, New Zealand. [Online]. Available: http://globalasthmareport.org/2014/Global_Asthma_Report_2014.pdf

[2] Forum of International Respiratory Societies. (2017). *The Global Impact of Respiratory Disease—Second Edition*. Sheffield. [Online]. Available: https://www.who.int/gard/publications/The_Global_Impact_of_Respiratory _Disease.pdf

[3] B. M. Rocha, D. Filos, L. Mendes, G. Serbes, S. Ulukaya, Y. P. Kahya, N. Jakovljevic, T. L. Turukalo, I. M. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jácome, A. Marques, N. Maglaveras, R. P. Paiva, I. Chouvarda, and P. de Carvalho, "An open access database for the evaluation of respiratory sound classification algorithms," *Physiolog. Meas.*, vol. 40, no. 3, Mar. 2019, Art. no. 035001.

[4] P. M. Shakeel, M. A. Burhanuddin, and M. I. Desa, "Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks," *Measurement*, vol. 145, pp. 702–712, Oct. 2019.

[5] Y. M. Arabi *et al.*, "How the COVID-19 pandemic will change the future of critical care," *Intensive Care Med.*, vol. 47, no. 3, pp. 282–291, 2021.

[6] J. Li, J. Yuan, H. Wang, S. Liu, Q. Guo, Y. Ma, Y. Li, L. Zhao, and G. Wang, "LungAttn: Advanced lung sound classification using attention mechanism with dual TQWT and triple STFT spectrogram," *Physiolog. Meas.*, vol. 42, no. 10, Oct. 2021, Art. no. 105006.

[7] X. Zhao, Y. Shao, J. Mai, A. Yin, and S. Xu, "Respiratory sound classification based on BiGRU-attention network with XGBoost," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2020, pp. 915–920.

[8] A. Kandaswamy, C. S. Kumar, R. P. Ramanathan, S. Jayaraman, and N. Malmurugan, "Neural classification of lung sounds using wavelet coefficients," *Comput. Biol. Med.*, vol. 34, no. 6, pp. 523–537, 2004.

[9] K. Kc, Z. Yin, M. Wu, and Z. Wu, "Evaluation of deep learning-based approaches for COVID-19 classification based on chest X-ray images," *Signal, Image Video Process.*, vol. 15, no. 5, pp. 959–966, Jul. 2021.

[10] V. Despotovic, M. Ismael, M. Cornil, R. M. Call, and G. Fagherazzi, "Detection of COVID-19 from voice, cough and breathing patterns: Dataset and preliminary results," *Comput. Biol. Med.*, vol. 138, Nov. 2021, Art. no. 104944.

[11] N. Soni, I. Saini, and B. Singh, "AFD and chaotic map-based integrated approach for ECG compression, steganography and encryption in e-healthcare paradigm," *IET Signal Process.*, vol. 15, no. 5, pp. 337–351, Jul. 2021.

[12] M. B. Shuvo, R. Ahommed, S. Reza, and M. M. A. Hashem, "CNL-UNet: A novel lightweight deep learning architecture for multimodal biomedical image segmentation with false output suppression," *Biomed. Signal Process. Control*, vol. 70, Sep. 2021, Art. no. 102959.

[13] A. Joshi, R. Kumar, and C. Tiwari, "Enhanced exploration of chronic cough using improved convolutional neural networks and remote monitoring harnessing Internet of Things (IoT)," *Mater. Today, Proc.*, vol. 46, pp. 6465–6473, Jan. 2021.

[14] Y. Alotaibi and A. F. Subahi, "New goal-oriented requirements extraction framework for e-health services: A case study of diagnostic testing during the COVID-19 outbreak," *Bus. Process Manage. J.*, vol. 28, no. 1, pp. 273–292, Feb. 2022.

[15] R. Naves, B. H. G. Barbosa, and D. D. Ferreira, "Classification of lung sounds using higher-order statistics: A divide-and-conquer approach," *Comput. Methods Programs Biomed.*, vol. 129, pp. 12–20, Jun. 2016.

[16] F.-S. Hsu, S.-R. Huang, C.-W. Huang, C.-J. Huang, Y.-R. Cheng, C.-C. Chen, J. Hsiao, C.-W. Chen, L.-C. Chen, Y.-C. Lai, B.-F. Hsu, N.-J. Lin, W.-L. Tsai, Y.-L. Wu, T.-L. Tseng, C.-T. Tseng, Y.-T. Chen, and F. Lai, "Benchmarking of eight recurrent neural network variants for breath phase and adventitious sound detection on a self-developed open-access lung sound database—HF_Lung_V1," *PLoS ONE*, vol. 16, no. 7, Jul. 2021, Art. no. e0254134.

[17] L. Shi, K. Du, C. Zhang, H. Ma, and W. Yan, "Lung sound recognition algorithm based on VGGish-BiGRU," *IEEE Access*, vol. 7, pp. 139438–139449, 2019.

[18] J. Acharya and A. Basu, "Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 3, pp. 535–544, Jun. 2020.

[19] E. Cakir, G. Parascandolo, T. Heittola, H. Huttunen, and T. Virtanen, "Convolutional recurrent neural networks for polyphonic sound event detection," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 6, pp. 1291–1303, Jun. 2017.

[20] N. Peng, A. Chen, G. Zhou, W. Chen, W. Zhang, J. Liu, and F. Ding, "Environment sound classification based on visual multi-feature fusion and GRU-AWS," *IEEE Access*, vol. 8, pp. 191100–191114, 2020.

[21] S. Jayalakshmy and G. F. Sudha, "Scalogram based prediction model for respiratory disorders using optimized convolutional neural networks," *Artif. Intell. Med.*, vol. 103, Mar. 2020, Art. no. 101809.

[22] F. Demir, A. M. Ismael, and A. Sengur, "Classification of lung sounds with CNN model using parallel pooling structure," *IEEE Access*, vol. 8, pp. 105376–105383, 2020.

[23] F. Demir, A. Sengur, and V. Bajaj, "Convolutional neural networks based efficient approach for classification of lung diseases," *Health Inf. Sci. Syst.*, vol. 8, no. 1, pp. 1–8, Dec. 2020.

[24] V. Basu and S. Rana, "Respiratory diseases recognition through respiratory sound with the help of deep neural network," in *Proc. 4th Int. Conf. Comput. Intell. Netw. (CINE)*, Feb. 2020, pp. 1–6.

[25] K. K. Lella and A. Pja, "Automatic COVID-19 disease diagnosis using 1D convolutional neural network and augmentation with human respiratory sound based on parameters: Cough, breath, and voice," *AIMS Public Health*, vol. 8, no. 2, p. 240, 2021.

[26] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[27] G. Suryanarayana, K. Chandran, O. I. Khalaf, Y. Alotaibi, A. Alsufyani, and S. A. Alghamdi, "Accurate magnetic resonance image super-resolution using deep networks and Gaussian filtering in the stationary wavelet domain," *IEEE Access*, vol. 9, pp. 71406–71417, 2021.

[28] F. Godin, J. Dambre, and W. De Neve, "Improving language modeling using densely connected recurrent neural networks," 2017, *arXiv:1707.06130*.

[29] A. Ullah, K. Muhammad, W. Ding, V. Palade, I. U. Haq, and S. W. Baik, "Efficient activity recognition using lightweight CNN and DS-GRU network for surveillance applications," *Appl. Soft Comput.*, vol. 103, May 2021, Art. no. 107102.

[30] S. Roy, I. Kiral-Kornek, and S. Harrer, "ChronoNet: A deep recurrent neural network for abnormal EEG identification," in *Proc. Conf. Artif. Intell. Med. Eur.* Cham, Switzerland: Springer, 2019, pp. 47–56.

[31] S. B. Shuvo, S. N. Ali, S. I. Swapnil, T. Hasan, and M. I. H. Bhuiyan, "A lightweight CNN model for detecting respiratory diseases from lung auscultation sounds using EMD-CWT-based hybrid scalogram," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 7, pp. 2595–2603, Jul. 2021.

[32] S.-Y. Jung, C.-H. Liao, Y.-S. Wu, S.-M. Yuan, and C.-T. Sun, "Efficiently classifying lung sounds through depthwise separable CNN models with fused STFT and MFCC features," *Diagnostics*, vol. 11, no. 4, p. 732, Apr. 2021.

[33] A. Ponomarchuk, I. Burenko, E. Malkin, I. Nazarov, V. Kokh, M. Avetisian, and L. Zhukov, "Project Achoo: A practical model and application for COVID-19 detection from recordings of breath, voice, and cough," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 2, pp. 175–187, Feb. 2022.

[34] T. Li, Y. Yin, K. Ma, S. Zhang, and M. Liu, "Lightweight end-to-end neural network model for automatic heart sound classification," *Information*, vol. 12, no. 2, p. 54, Jan. 2021.

[35] G. Altan, Y. Kutlu, Y. Garbi, A. Ö. Pekmezci, and S. Nural, "Multimedia respiratory database (RespiratoryDatabase@TR): Auscultation sounds and chest X-rays," *Natural Eng. Sci.*, vol. 2, no. 3, pp. 59–72, Oct. 2017.
[36] G.-C. Chang and Y.-F. Lai, "Performance evaluation and enhancement of lung sound recognition system in two real noisy environments," *Comput. Methods Programs Biomed.*, vol. 97, no. 2, pp. 141–150, Feb. 2010.
[37] A. Bohadana, G. Izbicki, and S. S. Kraman, "Fundamentals of lung auscultation," *New England J. Med.*, vol. 370, no. 8, pp. 744–751, Feb. 2014.
[38] J. A. Svoboda and R. C. Dorf, *Introduction to Electric Circuits*, 2nd ed. Hoboken, NJ, USA: Wiley, 2013. [Online]. Available: https://http://www.wiley.com
[39] A. Yadav, M. K. Dutta, and J. Prinosil, "Machine learning based automatic classification of respiratory signals using wavelet transform," in *Proc. 43rd Int. Conf. Telecommun. Signal Process. (TSP)*, Jul. 2020, pp. 545–549.
[40] P. Piirilä and A. R. A. Sovijärvi, "Crackles: Recording, analysis and clinical significance," *Eur. Respiratory J.*, vol. 8, no. 12, pp. 2139–2148, Dec. 1995.
[41] R. Palaniappan, K. Sundaraj, N. U. Ahamed, A. Arjunan, and S. Sundaraj, "Computer-based respiratory sound analysis: A systematic review," *IETE Tech. Rev.*, vol. 30, no. 3, pp. 248–256, 2013.
[42] A. M. Badshah, J. Ahmad, N. Rahim, and S. W. Baik, "Speech emotion recognition from spectrograms with deep convolutional neural network," in *Proc. Int. Conf. Platform Technol. Service (PlatCon)*, Feb. 2017, pp. 1–5.
[43] V. Tiwari, "MFCC and its applications in speaker recognition," *Int. J. Emerg. Technol.*, vol. 1, no. 1, pp. 19–22, 2010.
[44] F. Liu, T. Shen, Z. Luo, D. Zhao, and S. Guo, "Underwater target recognition using convolutional recurrent neural networks with 3-D Mel-spectrogram and data augmentation," *Appl. Acoust.*, vol. 178, Jul. 2021, Art. no. 107989.
[45] H. Meng, T. Yan, F. Yuan, and H. Wei, "Speech emotion recognition from 3D Log-Mel spectrograms with deep learning network," *IEEE Access*, vol. 7, pp. 125868–125881, 2019.
[46] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[47] D. Perna, "Convolutional neural networks learning from respiratory data," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2018, pp. 2109–2113.
[48] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
[49] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*.
[50] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[51] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[52] D. Berrar, "Cross-validation," in *Encyclopedia of Bioinformatics and Computational Biology*, vol. 1. Oxford, U.K.: Elsevier, 2019, pp. 542–545.
[53] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, Jan. 2014.
[54] S. U. Kumar and H. H. Inbarani, "Neighborhood rough set based ECG signal classification for diagnosis of cardiac diseases," *Soft Comput.*, vol. 21, no. 16, pp. 4721–4733, Aug. 2017.
[55] A. T. Azar and S. A. El-Said, "Performance analysis of support vector machines classifiers in breast cancer mammography recognition," *Neural Comput. Appl.*, vol. 24, no. 5, pp. 1163–1177, Apr. 2014.
[56] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, pp. 37–46, Apr. 1960.
[57] B. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme," *Biochim. Biophys. Acta-Protein Struct.*, vol. 405, no. 2, pp. 442–451, Oct. 1975.
[58] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, L. Yu, and Q. Ni, "A deep learning system to screen novel coronavirus disease 2019 pneumonia," *Engineering*, vol. 6, no. 10, pp. 1122–1129, Oct. 2020.
[59] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.

YOUNGJIN CHOI was born in Gumi, Gyeongbuk, South Korea, in 1990. He received the B.S. degree in biomedical science from Daegu University, South Korea, in 2015. He is currently pursuing an integrated Ph.D. degree with the Department of Industrial and Management Engineering, Korea University. His research interests include deep learning on sound processing, designing and applying deep learning in healthcare, and artificial intelligence.

HOERYEON CHOI received the B.S. degree in industrial engineering from Dankook University, South Korea, in 1993, and the Ph.D. degree in industrial engineering from Korea University, South Korea, in 2005. She is currently working as a Visiting Professor with the School of Industrial and Management Engineering, Korea University. Her research interests include big data analysis, machine learning, and AI algorithms (image generation and autonomous driving).

HWAYOUNG LEE received the B.S. degree in biological science from the Korean Advanced Institute of Science and Technology (KAIST), South Korea, in 2003, and the Ph.D. degree in internal medicine from The Catholic University of Korea, South Korea, in 2016. She is currently working as an Assistant Professor at Seoul St. Marys' Hospital, College of Medicine, The Catholic University of Korea. Her research interests include allergy, clinical immunology, and respiratory diseases.

SOOKYOUNG LEE received the B.S. degree in medicine and the Ph.D. degree in internal medicine from The Catholic University of Korea, South Korea, in 1988 and 1998, respectively. She is currently working as a Professor at Seoul St. Marys' Hospital, College of Medicine, The Catholic University of Korea. Her research interests include allergy, clinical immunology, and respiratory diseases.

HONGCHUL LEE received the B.S. degree in industrial engineering from Korea University, in 1983, the M.S. degree in industrial engineering from The University of Texas at Arlington, in 1988, and the Ph.D. degree in industrial engineering from Texas A&M University, in 1993. He is currently a Professor with the Department of Industrial Systems and Information Engineering, Korea University. His research interests include system engineering, system simulations, and artificial intelligence.