Journal of the Korean Institute of Industrial Engineers
Vol. 47, No. 2, pp. 208-216, April 2021.
ISSN 1225-0988 | EISSN 2234-6457
https://doi.org/10.7232/JKIIE.2021.47.2.208

이상 호흡음 분류를 위한 계층적 어텐션 네트워크 모델

정기원 · 김성범†

고려대학교 산업경영공학과

Abnormal Respiratory Sound Classification Using Hierarchical Attention Networks Model

Kee Won Jeong · Seoung Bum Kim

Department of Industrial and Management Engineering, Korea University

Recently the diagnosis of respiratory diseases based on respiratory-sound data has drawn much attention. However, no quantitative assessment method exists to detect abnormal respiratory sounds when making a diagnosis. Although existing studies have attempted to support doctors by providing them with machine learning results, there were limitations on explaining causal symptoms and establishing quantitative assessment. Therefore, a reliable method that can diagnose and explain the causal symptoms based on respiratory sound data is required. In this study, we propose using a hierarchical attention network for detecting abnormal respiratory sounds. The hierarchical attention network reflects hierarchical patterns of respiratory sounds and allows us to interpret the important feature of respiratory sounds. The experimental results showed that hierarchical attention network performed better than other existing methods in terms of sensitivity (ability of correctly detecting abnormal respiratory sounds) and explainability. We believe that the framework presented in this study can not only serve as a “second opinion” that can help doctors diagnose existing respiratory diseases, but also help doctors’ future research on unidentified respiratory diseases.

Keywords: Abnormal Respiratory Sound Classification, Computer Aided Diagnosis, Feature Extraction, Hierarchical Attention Networks, ICBHI 2017 Challenge

1. 서 론

2017년 기준 전 세계 사망 원인 10위 중 3위를 차지하고 있는 호흡기 질환은 조기에 발견하는 것이 중요하다고 알려져 있다. 현재 호흡기 질환은 의사의 청진을 통해 진단되고 있고, 의사는 환자의 호흡 주기 동안 발생하는 부잡음(adventitious sounds)을 이용해 호흡기 질환 종류를 판단한다. 부잡음은 호흡기 질환을 앓고 있는 환자의 호흡기류나 기도 주변의 폐 또는 흉막의 상태가 건강하지 않을 때 발생하는 소리로 대표적으로는 천명음(wheeze)과 수포음(crackle)이 있다. 천명음은 공기가 좁아진 기도 부위를 통과하면서 발생하는 연속적인 소리로 보통 100ms 이하 동안 지속하는 것으로 알려져 있다(Faustino, 2019). 천명음은 소리의 구성에 따라 단조성(monophonic)과 복조성(polyphonic)으로 분류된다. 단조성 천명음은 좁아진 기도에 의해 발생하는 단일 주파수 형태의 소리로 기관지결핵이나 기도협착증 환자에게서 발생하는 소리이다. 반면에 복조성 천명음은 기도의 내강이 좁아져서 발생하는 다중 주파수 형태의 소리로 만성폐쇄성폐질환 환자에게 들리는 소리이다(Faustino, 2019). 수포음은 공기가 액체로 차 있는 폐포 및 소기도를 통과하면서 발생하는 단속적인 소리로 보통 20ms 이하 동안 지속하는 것으로 알려져 있다. 수포음은 소리 크기에 따라 미세 수포음과 거친 수포음으로 나뉘며, 미세 수포음은 비교적 소리가 부드러우며 폐부종이나

제16회 석사논문경진대회 수상논문.
† 연락저자 : 김성범 교수, 02841 서울특별시 성북구 안암로 145, 고려대학교 산업경영공학과, Tel : 02-3290-3397, Fax : 02-929-5888, E-mail : sbkim1@korea.ac.kr
2021년 1월 18일 접수; 2021년 2월 25일 수정본 접수; 2021년 3월 4일 게재 확정.

만성폐쇄성폐질환 환자에게 발견되고, 거친 수포음은 미세 수포음보다 더 거칠고 낮은 음조의 형태의 소리로 소기도 기관지염 환자에게서 발생하는 소리이다(Faustino, 2019). 이처럼 실제 의료 현장에서는 의사가 환자의 호흡음에서 발생하는 소리를 청취하여 호흡기 질환을 진단하는 청진 진단 방식을 사용하고 있다. 그러나 청진에 의한 진단은 의사의 청력이나 진료 경험, 숙련도와 같은 주관적인 요인에 의존하기 때문에 진단 결과에 대한 신뢰성을 떨어트리고, 의사의 주관적인 요인에 따라서 오진 가능성을 높일 수가 있다. 최근 이러한 문제를 해결하기 위해 호흡음을 디지털화하고 정량적으로 분석하여 의사의 진단을 도울 수 있는 진단 지원 시스템(CAD : computer aided system) 개발을 위한 연구가 진행되고 있다.

컴퓨터와 같은 시스템이 호흡음을 이해하고 분석하기 위해서는 호흡음을 디지털화하고 특징을 추출하는 과정이 필요하다. Faustino(2019)는 이상 호흡음 분류를 위한 적절한 소리 특징 추출을 위해 호흡음 데이터로부터 power spectral density(PSD), mel spectrogram(MS), mel-frequency cepstral coefficients(MFCC) 세 가지 특징을 추출하고 특징마다 분류 모델을 학습 시켜서 특징 추출 방법론에 따른 호흡음 분류 성능을 비교하였다. Xie *et al*.(2012)는 호흡음의 특징 공간이 고차원이라는 점에 주목하여 특징 공간을 축소 시켜줄 수 있는 다중 스케일 주성분 분석 기반의 특징 추출 방법론을 제안하였다. 이후 다양한 호흡음 특징 추출 방법론을 기반으로 머신러닝 알고리즘을 적용한 이상 호흡음 분류 모델 방법론에 관한 연구가 진행되었다. Jakovljević *et al*.(2017)은 hidden Markov 모델을 이용한 호흡음 분류 모델 방법론을 제안하였고, Chambres *et al*.(2018)은 부스팅 의사 결정 나무 모델을 이용한 분류 모델 방법론을 제안하여 기존 방법론들보다 우수한 호흡음 분류 성능을 보였다. 그러나 머신러닝 알고리즘을 적용한 방법론들은 호흡음으로부터 학습 모델에 적합한 특징을 추출하기 위해 많은 시간과 노력이 필요했고, 대부분의 연구가 호흡음이 정상인지 비정상인지 이진 분류 문제를 풀었다는 점에서 실제 의료 현장에 적용되기에는 다소 한계점이 있다. 최근에는 이러한 한계점을 개선하기 위해 모델 스스로 호흡음의 특징을 학습하는 딥러닝 알고리즘 기반 방법론이 제안되고 있다. Ma *et al*.(2019)는 호흡음 특징으로 스펙트로그램(spectrogram)과 웨이블릿 분석 행렬(wavelet analysis matrix) 두 가지를 사용한 ResNet 모델 기반의 호흡음 분류 방법론을 제안하였고, Minami *et al*.(2019)은 스펙트로그램과 스칼로그램(scalogram)을 사용하여 VGG16 모델 기반의 호흡음 분류 방법론을 제안하여 기존 머신러닝 알고리즘 보다 우수한 분류 성능을 보여주었다.

최근 호흡음을 분석하기 위한 딥러닝 알고리즘 기반의 방법론들이 제안되면서 기존 머신러닝 알고리즘 기반 방법론들의 한계점을 극복하고 의사의 진단을 도울 수 있는 인공지능 시스템이 가까운 미래에 실제 의료 현장에 사용될 것으로 기대되고 있다. 그러나 기존의 제안 방법론들은 입력된 호흡음에서 발생한 부잡음의 종류에 대해서만 예측하였다. 실제 의료 현장에서는 의사가 부잡음 종류뿐만 아니라 소리 크기, 등장 시기와 같은 소리 정보도 함께 이용해서 호흡기 질환을 진단하게 되는데 단순히 의사에게 부잡음 종류만 알려줄 수 있는 시스템은 실제 의료 현장에서 의사 진단에 도움을 줄 수 있는 용도로 사용되기에는 한계가 있다. 따라서 본 연구에서는 호흡음의 중요 시간과 주파수 영역을 계층적으로 탐색하는 과정을 통해서 호흡음 분류 성능을 향상시킬 뿐만 아니라, 이상 호흡음에 대한 중요 소리 정보도 함께 제공할 수 있는 방법론을 제안한다. 제안한 방법론은 international conference on biomedical and health informatics (ICBHI)에서 공개한 대규모 호흡음 데이터셋(Rocha, 2017)에 적용해 다른 방법론들보다 더 나은 분류 성능을 보임을 확인하였다. 본 논문의 중요 기여점은 다음과 같다.

- 호흡음의 구조적 특성을 반영해서 중요 시간과 주파수 영역을 계층적으로 탐색하는 딥러닝 모델 기반 방법론을 제안한다.
- 계층적 어텐션 네트워크에서 산출된 어텐션 스코어를 통해 호흡기 질환 진단에 도움이 될 수 있는 정보를 시각적으로 제공한다.

본 논문은 다음과 같은 구조를 가진다. 제 2장에서는 ICBHI 2017 Challenge에서 공개한 호흡음 분류를 위한 머신러닝 알고리즘 기반 방법론과 최근에 연구된 딥러닝 알고리즘 기반 방법론을 설명한다. 제 3장에서는 본 연구에서 제안하는 데이터 전처리 및 특징 추출 알고리즘과 계층적 어텐션 네트워크를 활용한 호흡음 분류 방법론에 대하여 설명한다. 제 4장에서는 본 연구에서 사용된 데이터와 평가지표, 실험 내용과 관련된 하이퍼파라미터 설정에 대해 설명하고 제안하는 방법론의 성능을 실험을 통해 보인다. 마지막으로 제 5장에서는 결론 및 추후 연구과제를 제시한다.

2. 관련 연구

본 연구는 호흡음에 대한 시간과 주파수 영역을 계층적으로 학습하여 중요 시간 및 주파수 구간을 탐색하고 이를 통해 이상 호흡음에 대한 소리 정보를 제공할 수 있는 방법론을 제안하고자 한다. 호흡음 분류 모델을 개발하고 모델 성능을 검증하기 위해서는 공개된 데이터베이스가 필요하다. 본 장에서는 2017년에 열렸던 대규모 데이터베이스 호흡음 분석 대회에서 가장 우수한 성능을 보였던 방법론에 관해 설명하고 이후 동일한 데이터로 연구된 딥러닝 알고리즘 기반 방법론에 관해 설명한다.

2.1 ICBHI 2017 Challenge

지난 수십 년 동안 호흡음 분석에 대한 연구가 꾸준하게 진행되었지만, 공개된 대규모 데이터베이스가 구축되어 있지 않아

방법론 간 객관적인 성능 비교가 어려웠다. 이에 2017년 ICBHI Challenge는 대규모 호흡음 데이터셋을 공개하였고, 다양한 호흡음 분석 알고리즘이 개발되는 계기가 되었다. 대회에서 공개한 데이터셋은 다양한 기관으로부터 수집되었기 때문에 샘플마다 호흡음이 측정된 신체 부위, 녹음 장비가 다르고 실제 의료현장에서 발생할 수 있는 노이즈가 존재하여 호흡음을 위한 분류 모델을 구축하는데 어려움을 주었다. 당시 대회에서 SUK(Serbes, Ulukaya, and Kahya) 팀이 ICBH에서 공개한 베이스라인 방법론의 성능과 비교했을 때 가장 우수한 분류 성능을 보여 우승하였다(Rocha, 2019). ICBHI에서 공개한 베이스라인 방법론은 모든 호흡음 데이터 샘플링 비율(Sampling rate)을 4000Hz로 설정하고 MFCC 알고리즘을 통해 추출한 호흡음 특징을 의사결정나무 모델로 학습하였다. 의사결정나무의 분류 정확도는 정상 호흡음과 이상 호흡음에 대해 각각 75%와 12% 성능을 보였다. SUK 팀은 베이스라인 방법론과 동일하게 모든 호흡음 데이터에 대해 샘플링 비율을 4,000Hz로 설정하고, 분류 모델에 사용할 적절한 특징 추출을 위해 크게 3가지 과정을 거쳤다. 첫 번째 과정으로 호흡음 이외에 기침 소리, 심장 박동음과 같은 노이즈를 제거하기 위해 샘플 데이터마다 12th order band-pass 필터링을 적용했다. 두 번째 과정으로 이상 호흡음 중 천명음은 고주파 영역에서 발견되고 수포음은 저주파 영역에서 발견되기 때문에 tunable Q-factor 웨이블릿 변환을 이용하여 노이즈가 제거된 호흡음을 고주파, 저주파, 나머지 주파수 영역으로 분해하였다. 마지막 세 번째 과정에서는 각 주파수 영역마다 단시간 푸리에 변환(short time fourier transform) 알고리즘을 이용해 스펙트로그램을 추출하고, tunable Q-factor wavelet 알고리즘을 통해 웨이블릿 계수를 추출하였다. 추출된 스펙트로그램, 웨이블릿 계수는 고차원 공간으로 이루어져 있기 때문에 각각 6가지 통계량(평균, 표준편차, 최솟값, 최댓값, 왜도, 첨도)으로 요약하여 최종 호흡음의 특징으로 사용하였다. 분류 모델은 support vector machine (SVM) 모델을 사용했고, 정상 호흡음에 대한 분류 정확도 78%, 이상 호흡음에 대한 분류 정확도 20%로 ICBHI에서 제안한 베이스라인 방법론과 비교해 더 나은 분류 성능을 보였다.

2.2 Deep Learning Algorithm for Respiratory Sound

SUK 팀이 제안한 방법론과 같이 머신러닝 알고리즘을 사용한 방법론들은 분류 모델에 적합한 특징을 추출하기 위해 전문성과 시간이 필요하다. 최근에는 호흡음으로부터 특징을 추출하는 과정에 최대한 사람이 개입하지 않고 스스로 특징을 학습할 수 있는 딥러닝 방법론들이 제시되고 있다. Minami *et al*.(2019)는 호흡음 샘플마다 단시간 푸리에 변환 알고리즘과 웨이블릿 변환 알고리즘을 적용해 스펙트로그램과 스칼로그램 특징을 추출하고 두 특징을 이미지로 변환하여 사용했다. 분류 모델은 이미지넷(ImageNet) 데이터를 사전 학습한 VGG-16 모델로 미세 조정(fine tuning)하여 학습시켰다. 두 특징 이미지마다 VGG16 모델을 독립적으로 학습시켜서 각 이미지로부터 특징을 추출하게 되고 추출된 특징을 결합한 후 최종적으로 완전 연결 레이어(fully-connected layer)을 통과 시켜 호흡음의 클래스를 분류하였다. Minami *et al*.(2019)이 제안한 방법론의 경우, 정상 호흡음 분류 정확도 81%, 이상 호흡음 분류 정확도 28%로 기존 머신러닝 방법론들과 비교해 향상된 성능을 보여주었다. Ma *et al*.(2019)는 호흡음 샘플마다 스펙트로그램과 웨이블릿 행렬 특징을 추출하고 두 특징을 이미지로 변환하여 사용했다. 각 이미지마다 독립적으로 ResNet모델을 학습시켰고, Minami *et al*.(2019) 방법론과 동일하게 ResNet모델로부터 추출된 특징을 결합한 후 완전 연결 레이어를 통과 시켜 호흡음 클래스를 분류하였다. Ma *et al*.(2019)이 제안한 방법론의 경우 정상 호흡음 정확도 69.2%, 이상 호흡음 정확도 31.12% 성능을 보이며 Minami *et al*.(2019)이 제안한 방법론보다 이상 호흡음의 정확도가 더 높은 것을 확인할 수 있었다. 이처럼 딥러닝 방법론은 호흡음에서 추출한 특징을 이미지로 변환하기 때문에 기존 머신러닝 방법론들이 필요로 했던 전처리 및 특징 추출의 복잡한 과정을 생략할 수 있었다.

3. 제안 방법론

본 장에서는 본 연구에서 사용한 호흡음 데이터 전처리 및 특징 추출 알고리즘과 본 연구에서 제안하는 계층적 어텐션 네트워크 모델을 이용해 시간과 주파수 영역을 계층적으로 학습하여 중요 구간을 탐색할 수 있는 방법론을 구체적으로 설명한다.

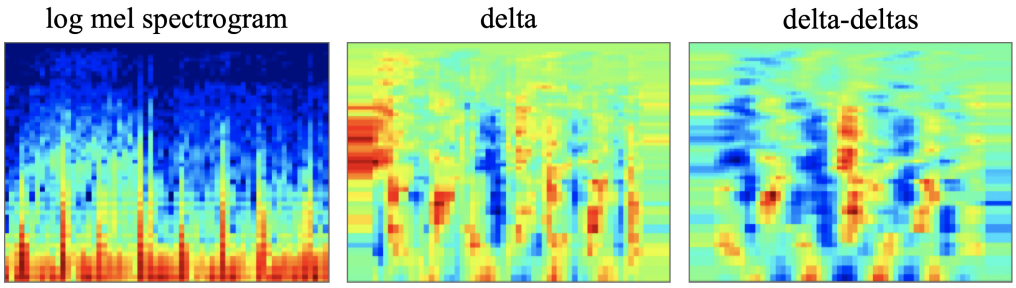
3.1. Data Preprocessing & Feature Extraction

본 연구에서 사용한 ICBHI 2017 데이터셋은 샘플마다 샘플링 비율이 다르기 때문에 가장 작은 샘플링 비율을 기준으로 모든 샘플에 대해 샘플링 비율을 4,000Hz로 설정하고, 이후 샘플 마다 단시간 푸리에 변환 알고리즘을 적용해 스펙트로그램을 추출하였다. 단시간 푸리에 변환 $F(n,w)$은 이동 구간 n과 시간 t, 윈도우 함수 $w(t)$, 변환 전 신호 값 $x(t)$, 허수 j, 주파수 ω에 의해 식 (1)로 표현된다. 윈도우 함수 $w(t)$은 hann 윈도우 함수로 식 (2)로 표현된다. 이후 사람의 청각 주파수 영역을 반영한 멜 스케일 필터(mel scale filter)와 로그 변환을 스펙트로그램에 적용해 로그 멜 스펙트로그램을 생성하였다. 멜 스케일 필터는 식 (3)으로 표현되고 f는 기존 주파수, m은 멜 단위 주파수, k는 헤르츠 단위 주파를 의미한다.

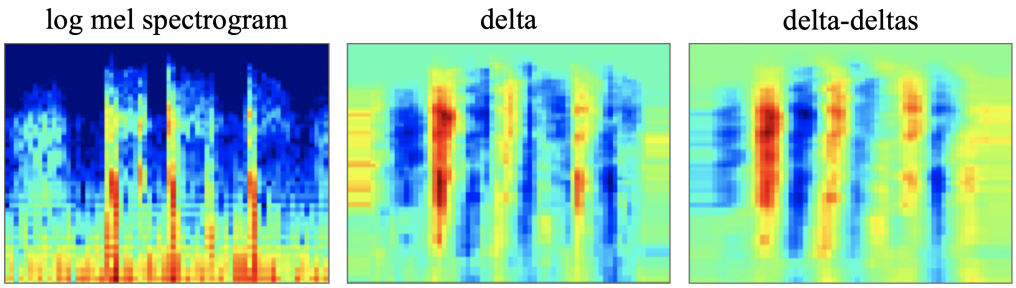
$$F(n,w)=\sum_{t=-\infty}^{\infty}x(t)w(t-n)e^{-j\omega t} \tag{1}$$

$$w(t)=0.54-0.46\cos\left(\frac{2\pi t}{L}\right) \tag{2}$$

$$H_m(k)=\begin{cases}0, & k<f(m-1)\\ \dfrac{k-f(m-1)}{f(m)-f(m-1)}, & f(m-1)\le k\le f(m)\\ 1, & k=f(m)\\ \dfrac{f(m-1)-k}{f(m+1)-f(m)}, & f(m)<k\le f(m+1)\\ 0, & k>f(m+1)\end{cases} \tag{3}$$



(a) Normal class



(b) Wheeze class

Figure 1. Features of Sample

추가적으로 로그 멜 스펙트로그램에 대하여 시간대별 차이 값을 구하는 델타(delta)와 델타에 대한 시간대별 차이 값을 구하는 델타-델타스(delta-deltas)를 계산하여 최종적으로 하나의 샘플에서 로그 멜 스펙트로그램, 델타, 델타-델타스 세 가지 특징을 추출하고 이들을 채널 축으로 쌓아 3차원 형태의 특징을 사용하였다. 로그 멜 스펙트로그램과 델타, 델타-델타스 특징에는 시간에 따라 변화하는 주파수 정보가 포함되어있기 때문에 소리의 시변적인(time variant) 특성을 반영할 수 있다. 본 연구에서는 L은 60(ms), n은 30(ms), 멜 스케일 필터는 64개를 적용하여 특징 추출하였다. <Figure 1>은 정상 호흡음과 이상 호흡음 샘플에 대하여 추출된 특징을 시각화한 그림이고, 좌측부터 로그 멜 스펙트로그램, 델타, 델타-델타스를 나타낸 것이다.

3.2 Hierarchical Attention Network based Feature of Sound Data Approach

본 연구에서는 호흡음의 시간과 주파수 영역을 계층적으로 학습하기 위해 계층적 어텐션 네트워크(hierarchical attention network; HAN)를 사용하였다. Yang *et al.*(2016)은 문서는 여러 개의 문장으로 구성되고 각 문장은 여러 개의 단어로 구성된 계층적 구조라는 점에 주목하여 이를 반영할 수 있는 계층적 어텐션 네트워크 알고리즘을 제안했다. 계층적 어텐션 네트워크 알고리즘은 단어 인코더, 단어 어텐션, 문장 인코더, 문장 어텐션, 분류기로 5개의 부분 네트워크로 이루어져 있으며, 단어 어텐션과 문장 어텐션은 문서의 클래스를 분류하는 데 중요한 역할을 하는 단어와 문장을 탐색해주는 역할을 한다. 단어 인코더와 문장 인코더에서는 bidirectional gated recurrent units(bidirectional GRU)를 사용한다. 일반적인 gated recurrent units(GRU)의 경우 특정 시점의 과거 정보만을 사용하는 반면에 bidirectional GRU는 특정 시점 이전의 과거 정보를 이용하는 정방향 GRU와 특정 시점 이후의 미래 정보를 이용하는 역방향 GRU로 구성되어 있기 때문에 입력된 단어 및 문장의 문맥적인 정보를 보다 더 잘 이해할 수 있도록 도와준다. 분류기는 중요 단어와 문장으로부터 산출된 값을 반영하여 소프트맥스 함수를 통해 문서를 분류한다. 계층적 어텐션 네트워크는 전통적인 문서 분류 알고리즘에 비해 높은 성능을 보였고, 어텐션 메커니즘을 이용해 문서 내 중요한 문장과 단어를 추론하는 목적으로도 유용하게 사용될 수 있다.

소리는 시간에 따른 진폭의 크기인 시간 영역과 주파수에 따른 진폭의 크기인 주파수 영역으로 구분할 수 있다. <Figure 2>는 소리를 표현할 수 있는 2가지 영역인 시간 영역(time domain)과 주파수 영역(frequency domain)을 나타낸다.

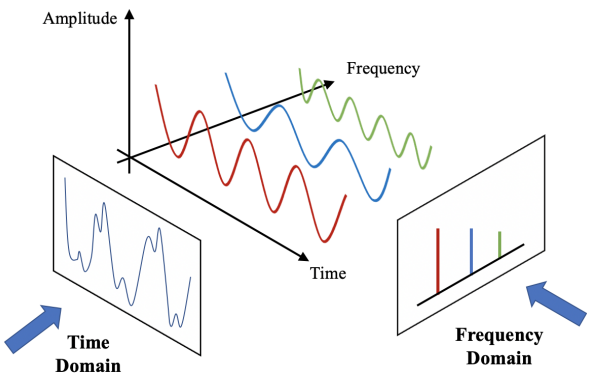


Figure 2. Time and Frequency Domains of Audio Signals

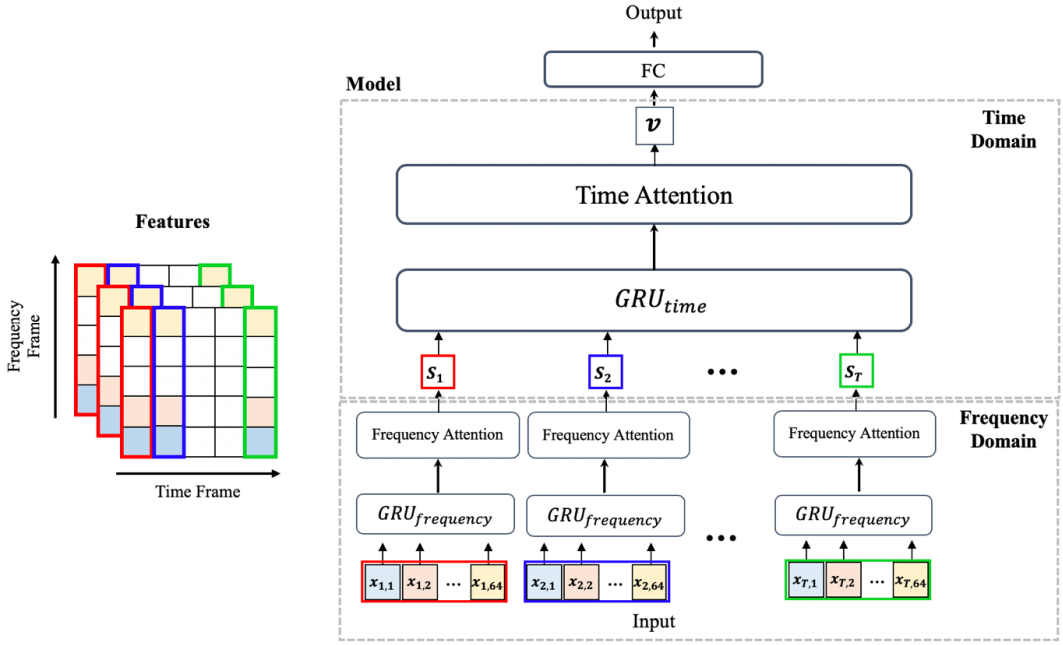


Figure 3. Proposed Hierarchical Attention Network for Respiratory Sound Feature

문서는 여러 개의 문장과 각 문장은 여러 개의 단어로 구성되어 있고, 소리는 여러 개의 시간 영역과 각 시간 영역은 여러 개의 주파수 영역으로 구성된다. 본 연구에서는 이러한 소리의 계층적인 구조를 반영하여 호흡음을 분석할 수 있는 계층적 어텐션 네트워크를 적용하였다. 본 연구에서 적용한 계층적 어텐션 네트워크의 학습 방식은 <Figure 3>에서 보여주고 있다. 주파수 영역과 시간 영역으로 구성되어 있는 네트워크를 아래와 같이 구체적으로 설명한다.

주파수 영역 : 모든 샘플 특징은 로그 멜 스펙트로그램, 델타, 델타-델타스를 3차원으로 표현한 $x_{t,f}$ 벡터로 정의할 수 있고, f는 주파수 프레임, t는 시간 프레임을 의미한다. 샘플 마다 64개의 멜 스케일 필터를 적용했기 때문에 모든 샘플의 주파수 프레임은 64개이고, 샘플 마다 길이가 다르므로 시간 프레임의 최대 길이를 T로 표현하였다. 따라서 모델의 입력은 $x_{t,f}$으로 식 (4)와 같이 표현된다. 주파수 영역 학습 과정에서는 시간대별 주파수 값을 어텐션 메커니즘이 적용된 bidirectional GRU에 입력하여 중요 주파수 정보를 반영한 히든 벡터(hidden vector)를 출력하도록 하였다. bidirectional GRU는 정방향 GRU와 역방향 GRU로 구성되어 있으며, $x_{t,f}$을 입력하면 식 (5)와 같이 정방향 히든 벡터($\overrightarrow{h}_{t,f}$)가 출력되고 식 (6)과 같이 역방향 히든 벡터($\overleftarrow{h}_{t,f}$)가 출력된다. 식 (7)과 같이 출력된 두 히든 벡터는 합쳐져 하나의 히든 벡터($h_{t,f}$)로 표현하였다.

$$input : x_{t,f},\ t = 1\ to\ T,\ f = 1\ to\ 64 \tag{4}$$

$$\overrightarrow{h}_{t,f} = \overrightarrow{GRU}_{frequency}(x_{t,f}),\ f = 1\ to\ 64 \tag{5}$$

$$\overleftarrow{h}_{t,f} = \overleftarrow{GRU}_{frequency}(x_{t,f}),\ f = 1\ to\ 64 \tag{6}$$

$$h_{t,f} = [\overrightarrow{h}_{t,f}, \overleftarrow{h}_{t,f}],\ f = 1\ to\ 64 \tag{7}$$

다음으로 중요 주파수 프레임을 찾기 위해 어텐션 메커니즘을 사용하였다. 우선 식 (8)과 같이 $h_{t,f}$을 one layer MLP에 통과 시켜 $h_{t,f}$의 히든 표현(hidden representation) $u_{t,f}$을 구하였다. $u_{t,f}$는 주파수 컨텍스트 벡터(context vector) u_{freq}와 내적 연산한 후, 식 (9)와 같이 소프트맥스 함수를 통해 어텐션 스코어 $\alpha_{t,f}$을 구할 수 있다. 주파수 컨텍스트 벡터 u_{freq}는 $u_{t,f}$와 동일한 크기의 파라미터이다. 해당 파라미터는 초기에는 난수 생성을 통해 임의값으로 설정되지만, 네트워크 내 주파수 영역이 학습하는 과정에서 함께 학습되는 파라미터로 학습이 진행됨에 따라 사용된 데이터의 전반적인 주파수 정보를 축약한다고 할 수 있다. 소프트맥스 함수를 통해 산출된 어텐션 스코어는 각 주파수 프레임의 중요도를 표현한 값이고, 64개 주파수 프레임 중요도의 총합은 1이 되게 된다. 최종적으로 어텐션 스코어 $\alpha_{t,f}$와 주파수 히든 벡터 $h_{t,f}$의 선형 결합 과정을 통해 중요 주파수 정보를 반영한 주파수 히든 벡터 s_t을 산출하였고, 이는 식 (10)으로 표현하였다.

$$u_{t,f} = \tanh(W_{freq} h_{t,f} + b_{freq}) \tag{8}$$

$$\alpha_{t,f} = \frac{\exp(u_{t,f}^T u_{freq})}{\sum_f \exp(u_{t,f}^T u_{freq})} \tag{9}$$

$$s_t = \sum_f \alpha_{t,f} h_{t,f} \tag{10}$$

시간 영역 : 주파수 영역에서 산출된 히든 벡터 s_t는 시간 영역 학습 과정에서 입력 값으로 사용되게 된다. s_t을 어텐션 메커니즘이 적용된 bidirectional GRU에 입력하여 중요 시간 정보를 반영한 히든 벡터를 출력할 수 있다. bidirectional GRU는 정방향 GRU와 역방향 GRU로 구성되어 있으며, s_t을 입력하면 식 (11)과 같이 정방향 히든 벡터($\overrightarrow{h}_t$)가 출력되고, 식 (12)와

같이 역방향 히든 벡터($\overleftarrow{h_t}$)가 출력된다. 식 (13)과 같이 출력된 두 히든 벡터를 합쳐 하나의 히든 벡터(h_t)로 표현하였다.

$$\overrightarrow{h_t} = \overrightarrow{GRU_{time}}(s_t),\ t=1\ to\ T \tag{11}$$

$$\overleftarrow{h_t} = \overleftarrow{GRU_{time}}(s_t),\ t=1\ to\ T \tag{12}$$

$$h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}],\ t=1\ to\ T \tag{13}$$

다음으로 중요 시간 프레임을 찾기 위해 어텐션 메커니즘을 사용하였다. 식 (14)와 같이 h_t를 one layer MLP에 통과시켜 h_t의 히든 표현 u_t을 구하고 시간 컨텍스트 벡터 u_{time}와 내적 연산한 후, 식 (15)와 같이 소프트맥스 함수를 통해 어텐션 스코어 α_t을 구하였다. 시간 컨텍스트 벡터 u_{time}는 u_t와 동일한 크기의 학습 파라미터로 주파수 컨텍스트 벡터 u_{freq}와 마찬가지로 네트워크 내 시간 영역이 학습하는 과정에서 함께 학습되는 파라미터이기 때문에 사용된 데이터의 전반적인 시간 정보를 축약한 파라미터라고 할 수 있다. 어텐션 스코어는 각 시간 프레임의 중요도를 표현한 값이고 T개 시간 프레임 어텐션 스코어 총합은 1이 된다. 이후 α_t와 히든 벡터 h_t의 선형 결합을 통해 중요 시간 정보를 반영한 히든 벡터 v를 산출하였고 이는 식 (16)으로 표현하였다. 시간 영역과 주파수 영역을 계층적으로 학습하여 산출된 히든 벡터 v는 입력된 호흡음의 중요 주파수와 시간 정보를 포함하고 있다.

$$u_t = \tanh(W_{time}h_t + b_{time}) \tag{14}$$

$$\alpha_t = \frac{ext(u_t^T u_{time})}{\sum_t \exp(u_t^T u_{time})} \tag{15}$$

$$v = \sum_t \alpha_t h_t \tag{16}$$

$$y_c = softmax(W_c v + b_c),\ c=1\ to\ 4 \tag{17}$$

$$\text{output}: y_c,\ c=1\ to\ 4 \tag{18}$$

최종적으로 히든 벡터 v는 식 (17)과 같이 완전 연결층 레이어를 거친 후 소프트맥스 함수를 통해 호흡음이 각 클래스에 속할 확률값을 계산하고, 가장 큰 확률값의 클래스로 호흡음을 분류한다. 따라서 모델의 최종적인 출력은 각 클래스의 확률값 y_c로 식 (18)과 같이 표현된다.

4. 실험 결과

본 장에서는 ICBHI 2017 Challenge 호흡음 데이터 및 평가 지표에 대한 설명과 실험 설정에 관련된 하이퍼파라미터 설정, 본 연구에서 제안하는 방법론과 비교 방법론과의 실험 결과를 포함하였다. 성능 비교를 위해서 support vector machine(SVM), random forest(RF), residual network(RN), ICBHI 2017 Challenge에서 공개한 방법론, 최근 제안된 딥러닝 기반 방법론인 Ma *et al.* (2019), Minami *et al.*(2019)을 사용하였다. 아울러 제안한 계층적 어텐션 네트워크를 통해 성능이 향상됨을 검증하기 위해 시간 영역에서만 어텐션 메커니즘을 적용한 학습 방식(single attention)과 시간과 주파수 영역 둘 다 어텐션 메커니즘을 적용한 학습 방식(hierarchical attention)의 분류 성능을 비교하였다. 마지막으로 어텐션 메커니즘에서 산출된 어텐션 스코어를 이용해 호흡기 질환을 진단하는데 중요한 정보를 시각적으로 제공할 수 있음을 보였다.

4.1 데이터 소개

본 연구에서 사용한 ICBHI Challenge 데이터 셋은 126명 환자의 호흡음을 녹음한 920개 소리 파일로 구성된다. 소리 파일은 여러 개의 호흡 주기로 이루어져 있고 전체 소리 파일의 호흡 주기 수는 6,898개이고, 호흡 주기마다 정상 호흡음, 천명음, 수포음, 천명음+수포음의 4개 클래스가 부여되었다. ICBHI Challenge에서는 훈련 데이터(training data) 4,142개, 테스트 데이터(testing data) 2,756개 인덱스 정보를 공개하였고, 훈련/테스트 샘플의 클래스 정보는 <Table 1>에서 구체적으로 보여주고 있다.

Table 1. Summary of the Training and Testing Sets

Class	Training set	Testing set	Total
Normal	2,063	1,579	3,642
Wheezes	501	385	886
Crackles	1,215	649	1,864
Wheezes+Crackles	363	143	506
Total	4,142	2,756	6,898

4.2 실험 설정

본 연구에서는 계층적 어텐션 네트워크 구축에 필요한 하이퍼파라미터로 주파수 영역의 bidirectional GRU 히든 차원 크기는 50, 시간 영역의 bidirectional GRU 히든 차원 크기는 100으로 설정하였다. 학습에 필요한 손실 함수는 분류 문제에 흔히 사용되는 교차 엔트로피(cross entropy)를 사용하였고, optimizer 함수로 AdamW 알고리즘을 사용하였다. 학습률(learning rate)은 0.001로 설정하였다.

4.3 평가 지표

본 연구에서는 ICBHI challenge에서 정의한 평가 지표를 사용하였다. 평가 지표는 정상 호흡음에 대한 정확도(S_p), 이상 호흡음에 대한 정확도(S_e), 정상/이상 호흡음 정확도의 산술평균(AS)과 조화 평균(HS) 총 4가지 평가 지표를 사용하였다. 평가 지표 내 인덱스는 <Table 2>에 구체적으로 표현하였다.

$$S_p = N_n / N \tag{19}$$

$$S_e = (C_c + W_w + B_b)/(C + W + B) \tag{20}$$

$$AS(Average\ Score) = (S_p + S_e)/2 \tag{21}$$

$$HS(Harmonic\ Score) = (2 \cdot S_p \cdot S_e)/(S_p + S_e) \tag{22}$$

Table 2. Determination Rules

		Prediction label			
		N*	W*	C*	(W+C)*
Reference label	N	N_n	N_w	N_c	N_b
	W	W_n	W_w	W_c	W_b
	C	C_n	C_n	C_w	C_c
	W+C	B_n	B_w	B_c	B_b

(N : Normal class, W : Wheezes class, C : Crackles class, W+C : Wheezes+ Crackles class)

4.4 실험 결과

(1) 분류 성능 평가 결과

<Table 3>은 본 연구에서 제안하는 방법론과 비교 대상이 되는 방법론의 분류 성능을 나타낸다.

Table 3. Comparison among algorithms in terms of SP, SE, AS, HS Using ICBHI 2017 Database. Boldface Values Represent the Best Performance for Each Measure

Algorithm	SP	SE	AS	HS
SVM	**86.38**	10.96	48.67	19.45
RF	82.33	14.52	48.42	24.69
ResNet18	67.26	26.08	46.67	37.59
ICBHI 2017 Baseline	75.00	12.00	43.00	15.00
SUK team	78.00	20.00	47.00	24.00
Ma *et al.*	69.20	31.12	50.16	42.93
Minami *et al.*	81.00	28.00	**54.00**	42.00
Proposed method (Single Attention)	65.99	29.39	47.69	40.66
Proposed method (Hierarchical Attention)	65.93	**34.75**	50.33	**45.51**

본 연구에서 제안하는 방법론의 경우 이상 호흡음에 대한 정확도(SE)가 34.75%로 다른 방법론들과 비교해 가장 우수한 성능을 보였다. 정상 호흡음에 대한 정확도(SP)는 본 연구에서 제안한 특징 추출 방법론으로 학습한 SVM이 86.38%로 가장 높지만, 이상 호흡음에 대한 정확도가 10.96%로 다른 방법론들과 비교해 가장 낮은 것을 확인할 수 있었다. 일반적으로 질병과 관련된 의료 분야에서는 음성과 양성 판정을 비교할 때 양성에 대한 정확도를 더 중요시하기 때문에 비록 SP가 높다고 하더라도 현실에 적용하기는 어렵다. 정상 호흡음 정확도와 이상 호흡음 정확도의 산술 평균(AS)의 경우 Minami *et al.* 알고리즘이 가장 좋은 성능을 보이지만 산술 평균은 특정 클래스 정확도 성능에 치우쳐 왜곡된 결과를 보일 수 있다. 이러한 한계점을 극복할 수 있는 지표가 조화 평균(HS)이고 본 연구에서 제안하는 알고리즘이 45.51%로 다른 알고리즘에 비해 가장 높은 것을 확인할 수 있었다. 또한, 전반적인 평가 지표에 대해 주파수와 시간 영역 모두 어텐션 메커니즘을 사용하는 것이 시간 영역에만 어텐션 메커니즘을 사용하는 것보다 더 좋은 성능을 보이는 것을 알 수 있었다. 따라서 실험을 통해 본 연구에서 제안하는 방법론은 소리의 구조적 특성을 반영해 계층적 어텐션 네트워크 모델을 사용함으로써 다른 방법론들에 비해 이상 호흡음 분류에 대한 뛰어난 성능이 있음을 확인하였다. 아울러 특정 클래스에 치우치지 않고 정상과 이상 호흡음에 대해 균형 있게 분류할 수 있음을 확인하였다.

(2) 어텐션 스코어 시각화

어텐션 메커니즘에서 산출된 어텐션 스코어를 이용하면 의사에게 호흡기 질환을 진단하는데 중요한 정보를 제공할 수 있다. 본 연구에서는 어텐션 메커니즘을 통해 중요 주파수를 나타내는 어텐션 스코어 $\alpha_{t,f}$와 중요 시간을 나타내는 어텐션 스코어 α_t을 산출하였다. 두 어텐션 스코어의 연산을 통해 주파수와 시간 영역 전체에 대한 어텐션 스코어 $\alpha'_{t,f}$를 식 (23)과 같이 계산하였다.

$$\alpha'_{t,f} = \alpha_t \times \alpha_{t,f} \quad \forall t \in [1,\ T] \tag{23}$$

호흡기 질환을 진단하기 위해서는 이상 호흡음의 종류뿐만 아니라 이상 호흡음의 등장 시기, 소리의 높낮이와 같은 시간과 주파수 정보도 함께 사용된다. 예를 들어 호흡 주기 초반에 수포음이 발생하거나 여러 소리가 합쳐진 것처럼 들리는 복조성 천명음이 발생하면 의사는 만성폐쇄성질환을 의심한다(Fasutino, 2019). ICBHI 2017 데이터는 각 호흡음 샘플마다 해당 환자의 질병 종류와 부잡음이 발생한 시간이 기록되어 있다. <Figure 4>는 만성폐쇄성폐질환을 앓고 있는 두 환자의 샘플에서 추출한 로그 멜 스펙트로그램과 어텐션 스코어를 시각화한 것이다. 하단의 히트맵은 멜 스펙트로그램을 나타낸 것이고, 상단의 히트맵은 어텐션 스코어를 시각화 한 것으로 큰 값일 수록 짙은 색을 띤다. 그리고 좌측은 환자의 호흡음에 수포음이 발생한 경우이고, 우측은 환자의 호흡음에 천명음이 발생한 경우이다. 붉은 점선 구간은 실제로 부잡음이 등장한 시간을 의미하는데 어텐션 스코어 히트맵을 보면 부잡음이 발생한 구간에서 어텐션 스코어가 높게 산출된 것을 확인할 수 있다. 만성폐쇄성질환 환자의 천명음과 수포음 특성도 시각적으로 확인할 수 있는데 수포음 샘플에서 산출된 어텐션 스코어는 중요 시간 구간이 호흡 초반에 짧게 나타나고, 천명음 샘플에서 산출된 어텐션 스코어는 여러 주파수 영역에 걸쳐서 높게 산출되어 복조성 정보를 시각적으로 보여주고 있다.

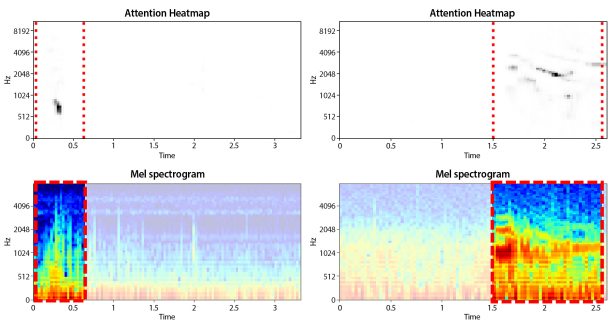


Figure 4. Attention Heatmap of Respiratory Sound Feature(Left : crackles / Right : wheeze)

<Figure 5>는 각 클래스 별 일부 샘플을 통해 산출한 어텐션 스코어와 멜 스펙트로그램을 시각화한 것이다. Correct는 모델의 정답 여부를 의미하며 예측한 클래스가 실제 클래스와 같으면 true, 틀리면 false로 표기하였다. 그림을 통해 모델이 클래스를 잘 못 예측했을 경우에도 부잡음이 일어난 구간에서 모델의 어텐션 스코어가 전반적으로 높게 나타나는 것을 확인할 수 있었다. 이러한 자료는 계층적 어텐션 네트워크의 어텐션 메커니즘이 중요 시간과 주파수에 대해서 적절하게 추론하고 있음을 보여준다.

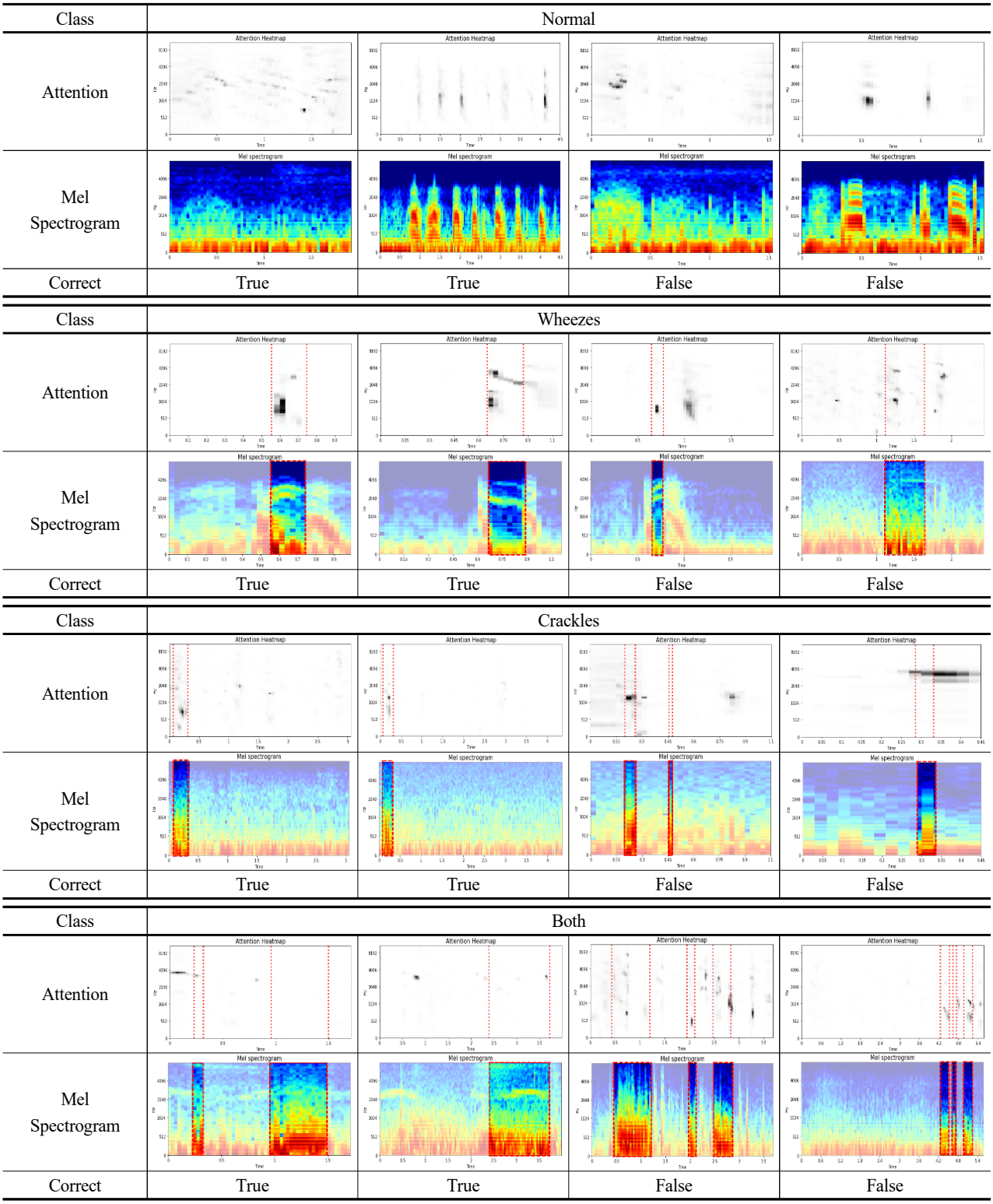


Figure 5. Example of Attention Heatmap

5. 결 론

의사는 청진을 통해 환자의 호흡 주기 동안 발생하는 이상 호흡음의 종류와 시간, 주파수 정보를 이용해서 질병을 판단하지만, 청진은 주관적인 판단으로 이루어지기 때문에 같은 환자에 대해서도 의사마다 다른 진단 결과가 나올 수도 있다. 이러한 문제를 보완하여 의사의 진단을 도울 수 있는 호흡음 분석 시스템에 관한 연구가 수행되고 있다. 연구 초기에는 머신러닝 알고리즘을 사용한 방법론들이 제안되었지만, 분류 모델에 적합한 특징 추출 및 학습을 위한 전문성과 많은 시간이 필요하다는 한계점이 있었다. 최근에는 딥러닝 방법론들이 제시되면서 사람의 개입을 최소화함과 동시에 향상된 분류 성능을 보였다. 하지만 최근 제안된 딥러닝 방법론들은 호흡음으로부터 추출된 특징을 이미지화하여 ResNet, VGG-16 모델과 같은 이미지 기반 학습 모델을 사용하고 있다. 제안된 방법론들은 모델이 주어진 호흡음의 특징을 스스로 학습하여 머신러닝 알고리즘과 비교해 우수한 분류 성능을 보이지만 호흡음의 시간에 따라 변하는 주파수 정보와 같이 소리의 시변적인 특성을 충분히 반영하지 못한다는 한계점이 있다. 본 연구에서는 소리의 구조적 특성을 반영하여 소리를 구성하는 시간과 주파수 영역을 순차적으로 학습할 수 있는 계층적 어텐션 네트워크를 적용한 호흡음 분류 방법론을 제안하였고, 실험을 통해 이미지 기반 모델의 딥러닝 방법론들보다 우수한 분류 성능을 보이는 것을 확인하였다. 또한 기존 방법론들은 호흡음의 클래스만을 예측하는 데 초점을 맞추었다. 하지만 실제 의료 현장에서는 이상 호흡음을 정확하게 분류하는 것과 더불어 이상 호흡음의 시간과 주파수 정보도 중요하게 사용된다. 본 연구에서는 계층적 어텐션 네트워크 모델의 어텐션 메커니즘을 이용하여 모델의 분류 성능 향상뿐만 아니라 어텐션 스코어를 시각화하여 이상 호흡음의 중요 시간과 주파수 정보를 제공할 수 있음을 실험을 통해 확인하였다.

본 연구는 문서 분류를 목적으로 개발된 계층적 어텐션 네트워크를 소리의 시간과 주파수 영역을 학습하기 위한 목적으로 사용한 첫 연구라는 점에서 의의가 있다. 본 연구에서 제안하는 방법론은 현재 증상을 명확하게 파악한 호흡기 질환들뿐만 아니라 새롭게 등장한 호흡기 질환의 중요 시간 및 주파수 특성을 발견하는데 큰 도움이 될 수 있을 것으로 본다. 또한 의료 현장 뿐만 아니라 특정 소리의 시간, 주파수 정보를 이용해 분류 문제를 해결해야 하는 다양한 산업 현장에서도 기여를 할 수 있을 것으로 기대한다.

참고문헌

Rocha *et al*. (2019), An Open Access Database for the Evaluation of Respiratory Sound Classification Algorithms, *Physiological Measurement*, **40**(3), 035001.

Breiman L. (2001), Random Forests, *Machine Learning*, **45**(1), 5-32.

Chambres, G., Hanna, P., and Desainte-Catherine, M. (2018), Automatic Detection of Patient with Respiratory Diseases Using Lung Sound Analysis, *In Proceedings-International Workshop on Content-Based Multimedia Indexing*.

Cortes, C. and Vapnik, V. (1995), Support Vector Networks, *Machine Learning*, **20**(3), 273-297.

Faustino, P. S. (2019), Crackle and Wheeze Detection in Lung Sound Signals Using Convolutional Neural Networks.

Jakovljević, N. and Lončar-Turukalo, T. (2018), Hidden Markov Model based Respiratory Sound Classification, In IFMBE Proceedings, *ICBHI : International Conference on Biomedical and Health Informatics*, Thessaloniki, Greece.

Liu *et al*. (2016), An Open Access Database for the Evaluation of Heart Sound Algorithms, *Physiological Measurement*, **37**(12), 2181-2213.

Ma, Y., Xu, X., Yu, Q., Zhang, Y., Li, Y., Zhao, J., and Wang, G. (2019), Lungbrn : A Smart Digital Stethoscope for Detecting Respiratory Disease Using Bi-Resnet Deep Learning Algorithm, *BioCAS 2019-Biomedical Circuits and Systems Conference*.

Minami, K., Lu, H., Kim, H., Mabu, S., Hirano, Y., and Kido, S. (2019), Automatic Classification of Large-Scale Respiratory Sound Dataset based on Convolutional Neural Network, *In 2019 19th International Conference on Control, Automation and Systems (ICCAS)*, IEEE, 804-807.

Palaniappan, R. and Sundaraj, K. (2013), Respiratory Sound Classification Using Cepstral Features and Support Vector Machine, *2013 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*.

Rocha *et al*. (2017), A Respiratory Sound Database for the Development of Automated Classification, *In International Conference on Biomedical and Health Informatics*, Springer, Singapore, 33-37.

Schonlau, M. and Zou, R. Y. (2020), The Random Forest Algorithm for Statistical Learning, *Stata Journal*, **20**(1), 3-29.

Xie, S., Jin, F., Krishnan, S., and Sattar, F. (2012), Signal Feature Extraction by Multi-Scale PCA and its Application to Respiratory Sound Classification, *Medical and Biological Engineering and Computing*, **50**, 759-768.

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., and Hovy, E. (2016), Hierarchical Attention Networks for Document Classification, *In 2016 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies*, San Diego, California, 1480-1489.

저자소개

정기원 : 한국외국어대학교 산업경영공학과에서 2020년 학사학위를 취득하고 고려대학교 산업경영공학과에서 석사 과정에 재학중이다. 연구분야는 Machine Learning, Deep Learning for signal data analysis이다.

김성범 : 한양대학교 산업공학과에서 1999년 학사를 취득하고 2001년과 2005년 미국 Georgia Institute of Technology에서 산업공학 석사학위, 박사학위를 취득하였다. 미국 텍사스 주립대학교 교수를 역임하고 2009년부터 고려대학교 산업경영공학부 교수로 재직하고 있다. 연구분야는 인공지능, 머신러닝, 최적화이다.