
Abnormal Respiratory Sound Classification Using Hierarchical Attention Networks

이상 호흡음 분류를 위한 계층적 어텐션 네트워크 모델

정기원, 김성범

고려대학교 산업경영공학과

목차

1. 프로젝트 개요
2. 프로젝트 주제
3. 데이터
4. 제안 방법론
5. 실험 결과
6. 결론

01 프로젝트 개요

❖ Abnormal Respiratory Sound (이상 호흡음)

- 부잡음(adventitious sounds) 라고도 불리며, 호흡기 질환을 앓는 환자에게 발생하는 소리
- 대표적인 이상 호흡음 종류로 천명음(Wheeze), 수포음(Crackle) 있음
- 이상 호흡음 마다 소리 형태나 지속시간, 발현 주파수와 같은 소리 특성이 다름

< 천명음, 수포음 차이점 >

	발생 이유	소리 형태	지속 시간	발현 주파수
천명음 (Wheeze)	좁아진 기도 부위를 공기가 통과하면서 발생	연속적인 소리	100ms 이하	100-2500Hz
수포음 (Crackle)	액체로 차 있는 폐포 및 소기도를 공기가 통과하면서 발생	단속적인 소리	20ms 이하	100-200Hz

Faustino, P. S. (2019). Crackle and wheeze detection in lung sound signals using convolutional neural networks.

01 프로젝트 개요

❖ 호흡기 질환 진단

- 의사는 청진을 통해 이상 호흡음 정보를 파악하고 이를 바탕으로 호흡기 질환을 진단
- 이상 호흡음의 소리 특성(소리 크기, 등장 시기, 주파수 형태 등)에 따라 호흡기 질환 종류가 다름

< 천명음, 수포음 종류에 따른 호흡기 질환 >



[주파수 형태에 따라]

천명음 — 단조성 천명음 ~ 기관지 결핵, 기도협착증
 — 복조성 천명음 ~ 만성폐쇄성폐질환

[소리 크기에 따라]

수포음 — 미세 수포음 ~ 폐부종, 만성폐쇄성폐질환
 — 거친 수포음 ~ 소기도 기관지염

✓ 청진에 의한 진단은 **의사의 주관적인 요인에 의존**하기 때문에 진단 결과에 대한 신뢰성을 떨어뜨리고, 오진 가능성을 높임

01 프로젝트 개요

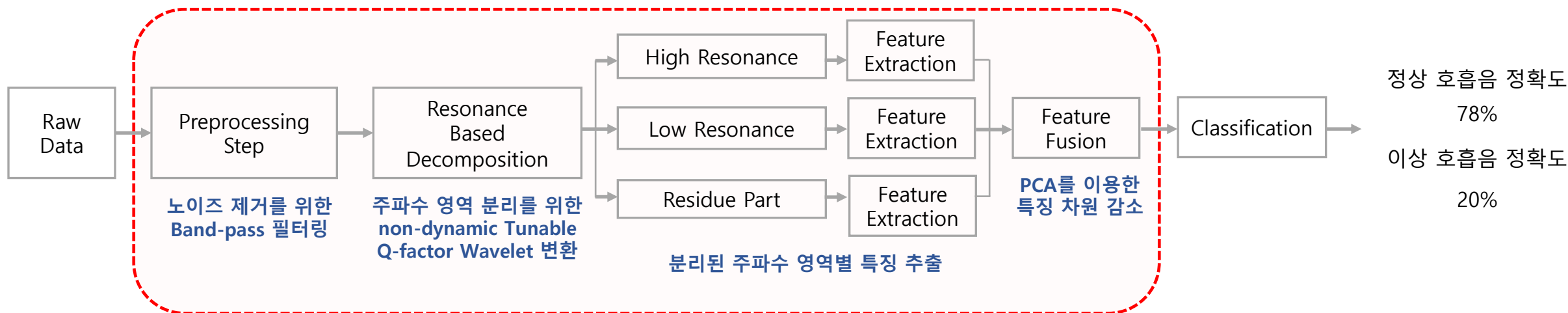
❖ 관련 연구

- 호흡음을 디지털화하고 정량적으로 분석하여 의사의 진단을 돕기 위한 연구가 활발히 진행
- 과거에는 적절한 소리 특징 추출 방법론과 머신 러닝 알고리즘 적용 방법론 제안

✓ ICBHI (International Conference on Biomedical and Health Informatics) 2017 Challenge

: 이상 호흡음 분석을 위한 알고리즘 개발을 목표로 개최된 대회.

< 우승팀 제안 방법론 구조 >



머신러닝 방법론 적용을 위한 복잡한 특징 추출 과정을 필요

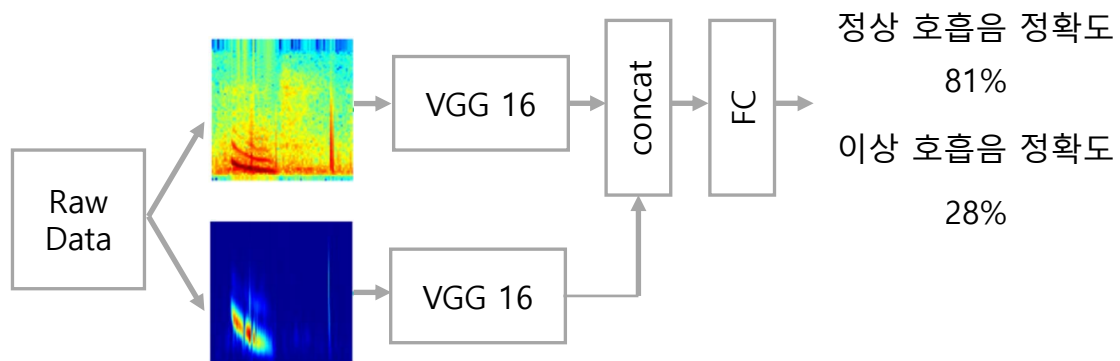
Liu, C., Springer, D., Li, Q., Moody, B., Juan, R. A., Chorro, F. J., ... Clifford, G. D. (2016). An open access database for the evaluation of heart sound algorithms. Physiological Measurement.

01 프로젝트 개요

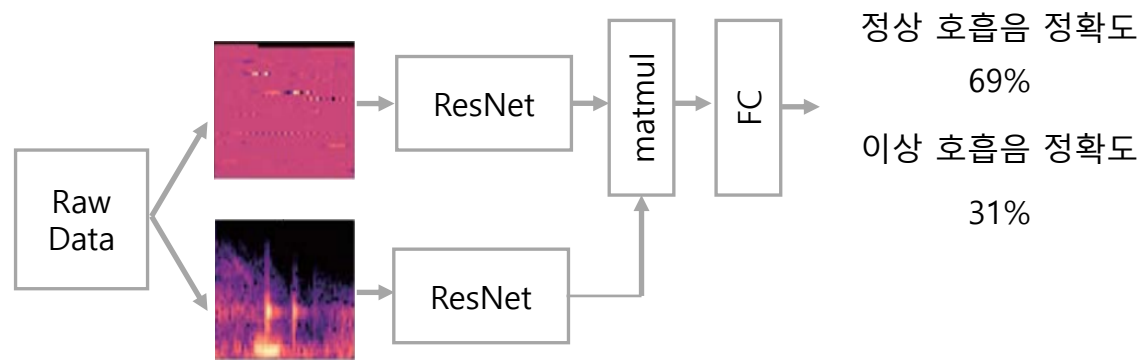
❖ 관련 연구

- 머신러닝 적용 방법론의 경우, 적합한 특징 추출을 위한 많은 시간과 노력을 필요
 - 최근에는 모델 스스로 호흡음 특징을 학습하는 딥러닝 기반 방법론 제안하며 이전 연구의 한계점을 극복
- ✓ 이미지 기반 학습의 딥러닝 모델 적용 방법론

< Minami et al. Method >



< Ma et al. Method >

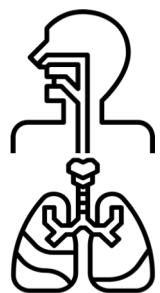


01 프로젝트 개요

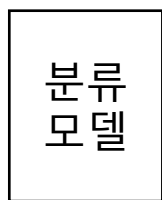
❖ 기존 연구의 한계점

- 이전 연구들은 입력된 호흡음에서 발생한 이상 호흡음의 종류에 대해서만 예측함.
- 실제 의료 현장에서는 이상 호흡음 종류 뿐만 아니라 소리 크기, 주파수와 같은 소리 정보도 함께 이용해서 호흡기 질환을 진단함.

< 기존 연구 결과 >



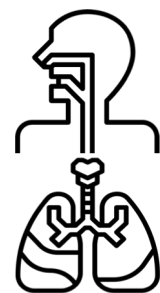
호흡음



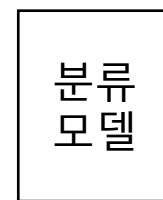
정상
천명음 발생
수포음 발생

“ 정상과 이상 호흡음 분류 ”

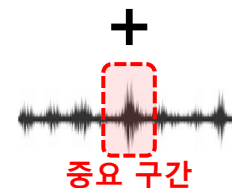
< 본 연구 제안 결과 >



호흡음



정상
천명음 발생
수포음 발생

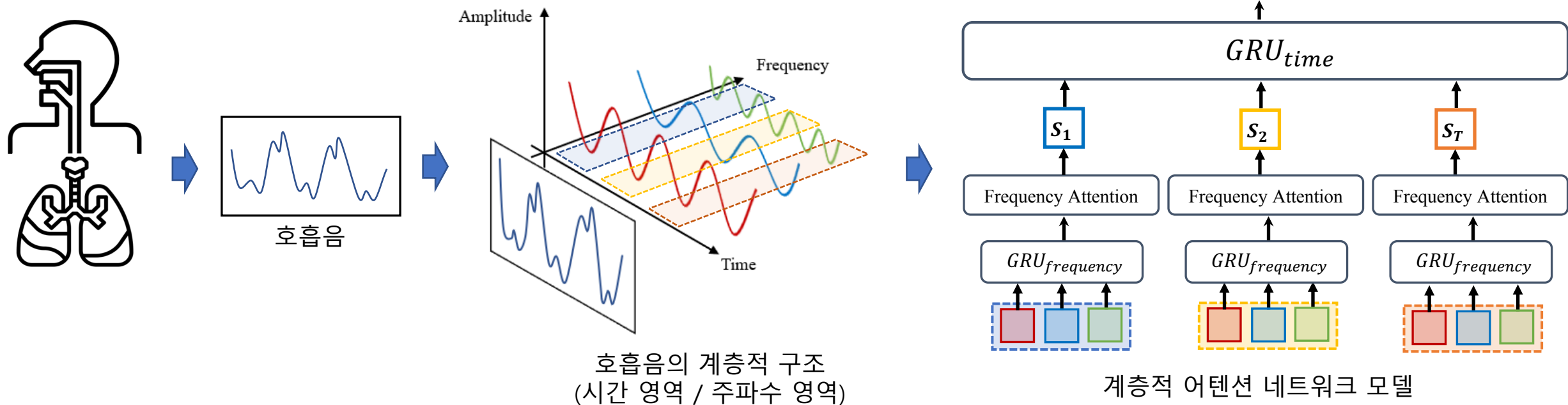


“ 정상과 이상 호흡음 분류 ”
+
“ 이상 호흡음 소리 정보 제공 ”

02 프로젝트 주제

❖ Abnormal Respiratory Sound Classification Using Hierarchical Attention Networks

- 호흡음의 구조적 특성을 반영해 **중요 시간과 주파수 영역을 계층적으로 탐색하는** 딥러닝 기반 방법론 제안
- 어텐션 메커니즘에서 산출된 어텐션 스코어를 통해 **이상 호흡음의 중요 소리 정보를 시각적으로 제공**



03 데이터

❖ ICBHI 2017 Challenge Data



ICBHI Challenge

International Conference on Biomedical and Health Informatics

- 호흡음 분석을 위해 공개된 대규모 데이터셋
- 126명 환자를 대상으로 호흡음을 녹음
 - 920개의 wav 파일과 920개의 text 파일로 구성
 - 각 소리 파일에는 여러 번의 호흡 주기로 구성 → 총 6898회 호흡주기
 - 실제 상황을 고려해 여러 노이즈가 섞여 있음
- 총 4개의 클래스로 구성
 - 총 6898회 中 **정상음** : 3638 / **천명음** : 885 / **수포음** : 1860 / **천명음+수포음** : 506

wav file

- 총 5.5시간
- 각 소리 파일은 10~70초
- 파일마다 호흡 주기는 평균 4회

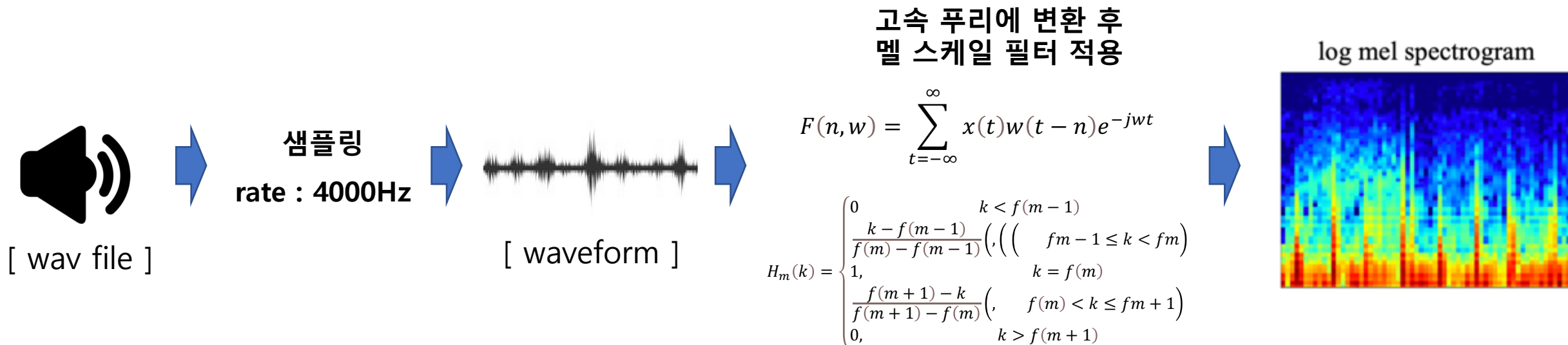
text file

- wav 파일의 주석
- 해당 호흡 주기의 클래스
- 해당 환자의 메타 정보 (호흡기 질환, 나이, 성별 등)

04 제안 방법론

❖ 전처리 및 특징 추출 방법

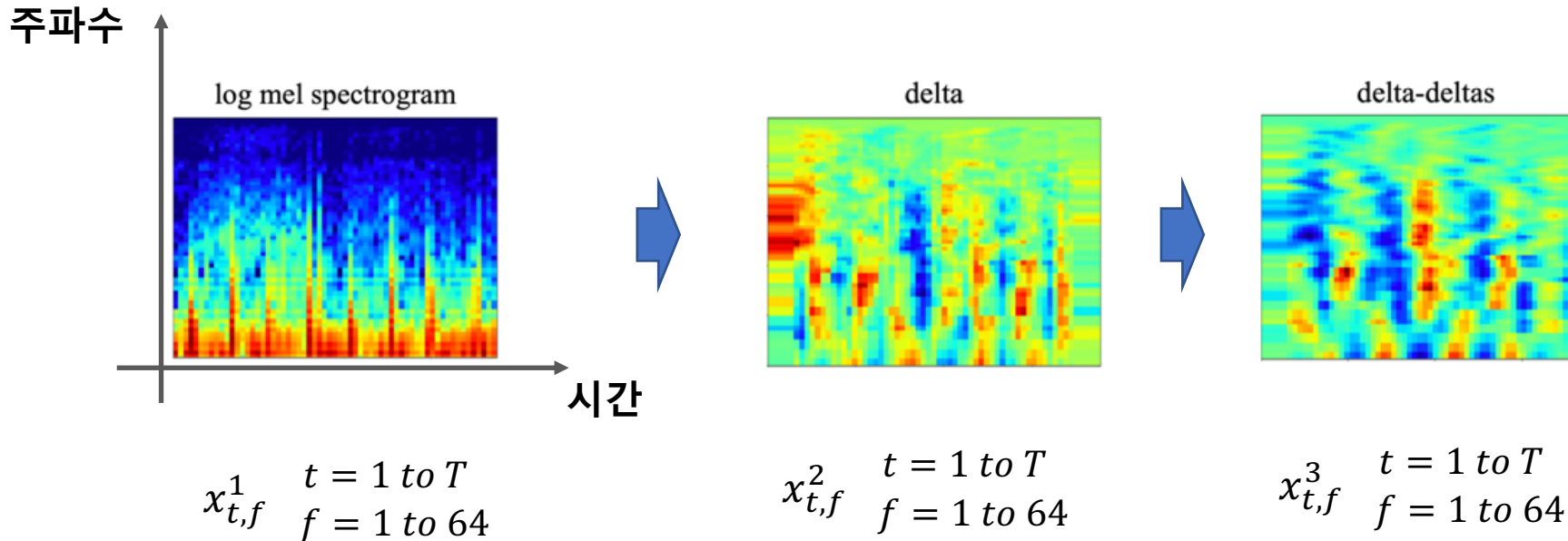
- 소리 파일 마다 sampling rate가 다르기 때문에 가장 낮은 4000Hz를 기준으로 샘플링
 - sampling rate : 이산적인 신호를 만들기 위해 연속적 신호에서 얻어진 단위 시간(초) 당 샘플링 횟수, 단위는 헤르츠를 사용.
 - 만약 sampling rate가 4000Hz라면 1초 당 4000개의 실수 값을 가짐
- 고속 푸리에 변환 알고리즘을 적용해 스펙트로그램을 추출
- 이후 멜 스케일 필터와 로그 변환을 통해 로그 멜 스펙트로그램 생성



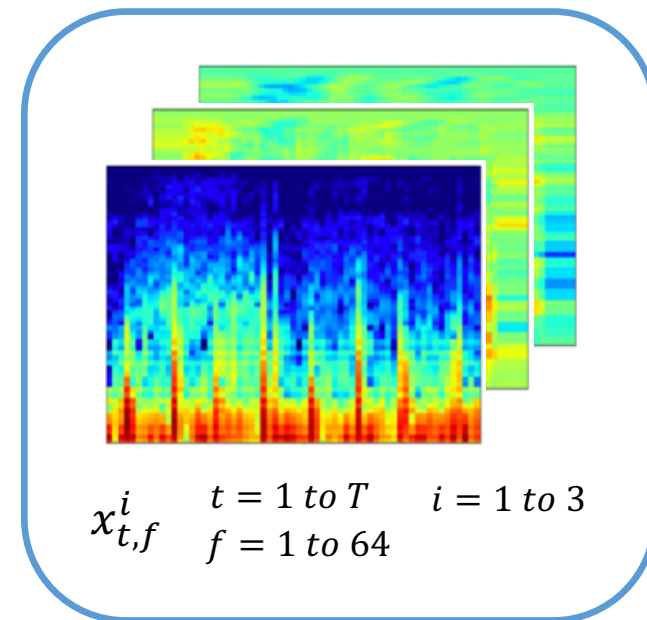
04 제안 방법론

❖ 전처리 및 특징 추출 방법

- 로그 멜 스펙트로그램은 시간과 주파수 영역을 가짐
- 이후 로그 멜 스펙트로그램으로부터 델타와 델타-델타스를 생성
 - 델타 특징 추출: 특징 간 상관 관계를 분리해주는 역할
- 3가지 특징을 채널 축으로 쌓아 3차원 형태의 특징으로 사용



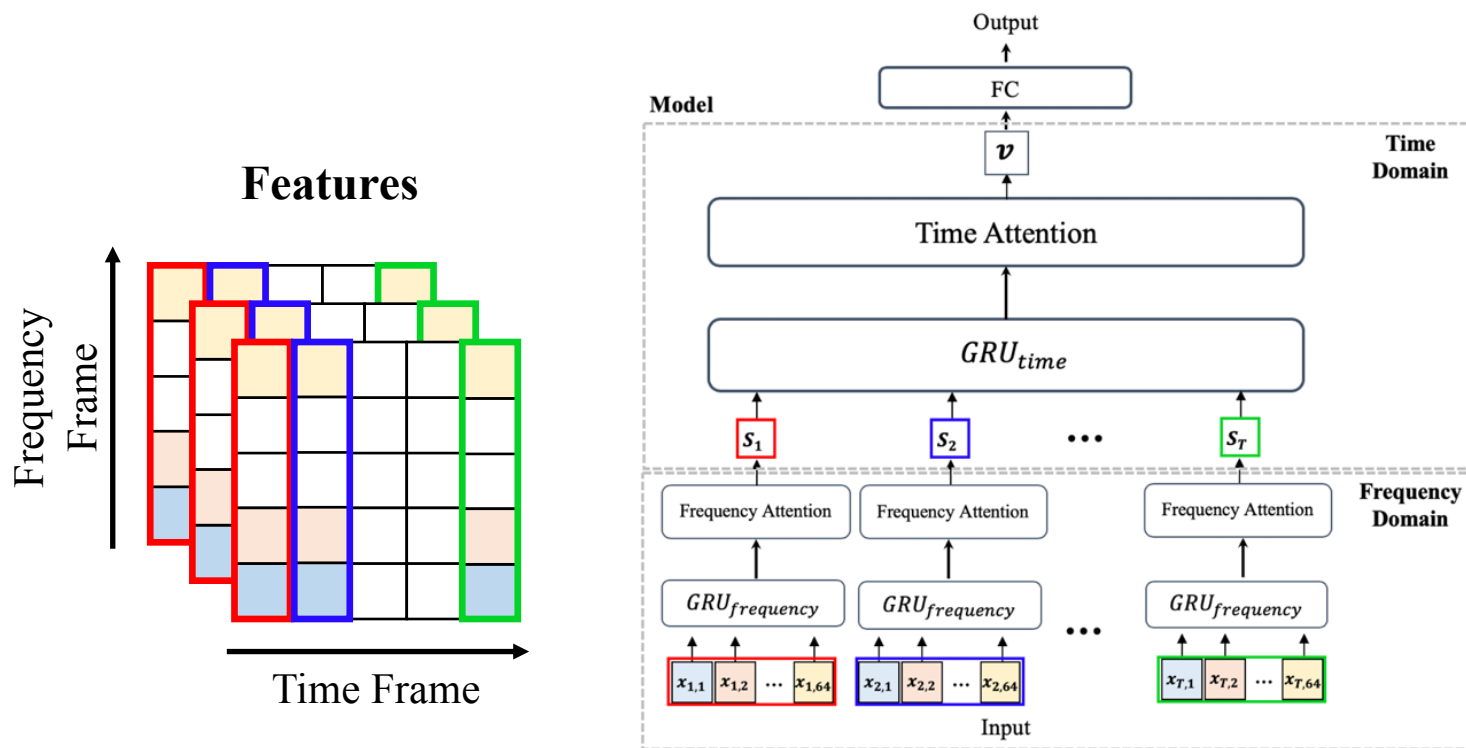
모델에 사용되는 Features



04 제안 방법론

❖ 제안하는 모델링 방법

- 추출한 특징은 여러 개 시간 구간과 각 시간 별 여러 주파수 구간으로 구성
- 각 시간 별 주파수 구간을 학습한 후 시간 구간을 학습하는 계층적 학습 방식을 제안



- ① 주파수 구간 학습
- ② 시간 구간 학습
- ③ 분류기

04 제안 방법론

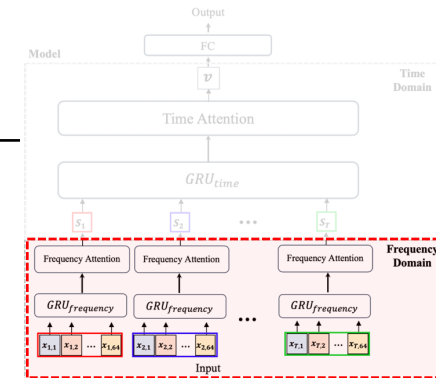
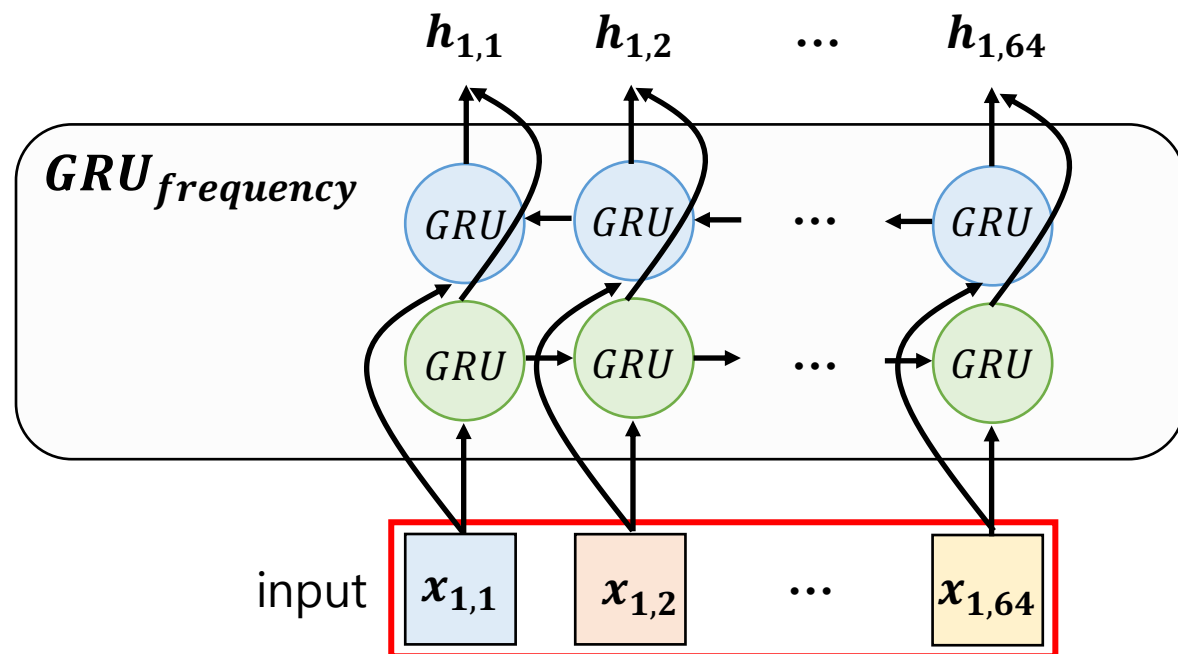
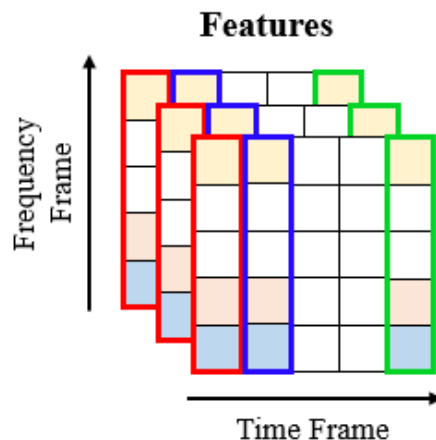
❖ 모델링 - ① 주파수 구간 학습

- 추출한 특징을 **Bidirectional GRU**에 입력
- 정방향과 양방향 히든 벡터가 합쳐진 하나의 히든 벡터 출력 $h_{t,f}$

$$\vec{h}_{t,f} = GRU_{frequency}(x_{t,f})$$

$$\overleftarrow{h}_{t,f} = GRU_{frequency}(x_{t,f})$$

$$h_{t,f} = [\vec{h}_{t,f}, \overleftarrow{h}_{t,f}]$$



04 제안 방법론

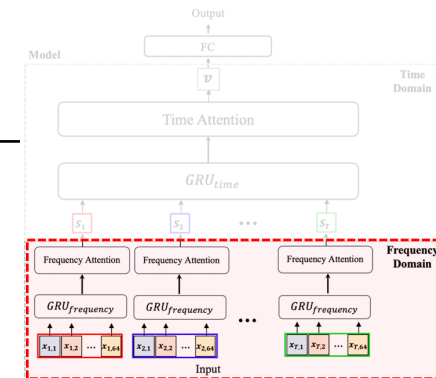
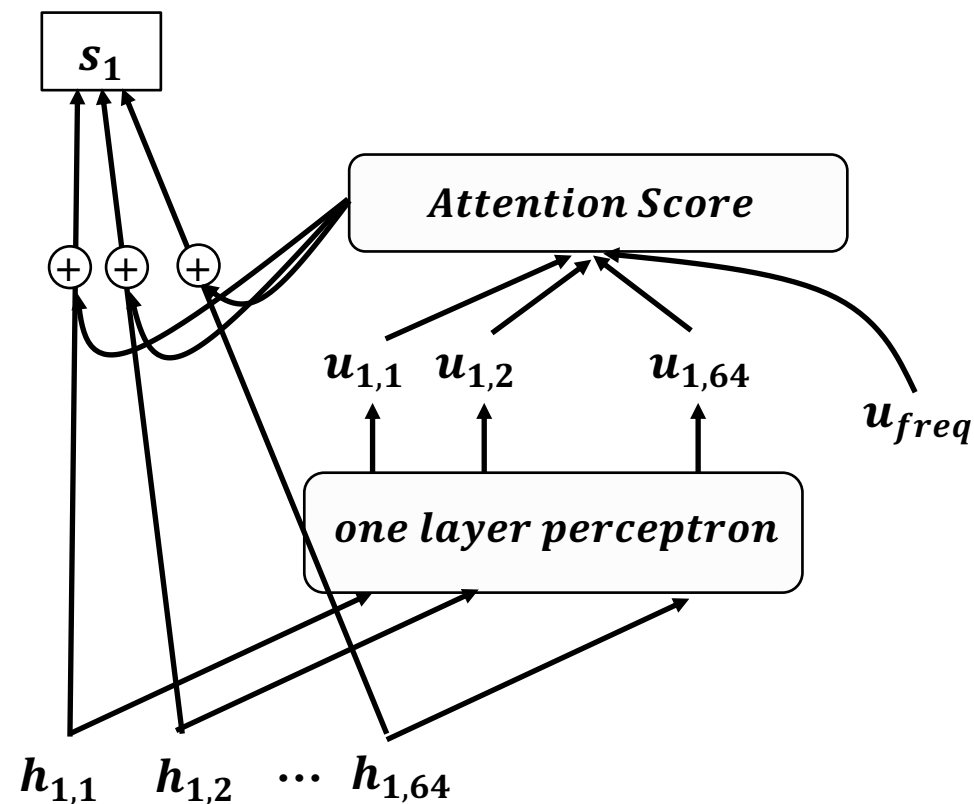
❖ 모델링 - ① 주파수 구간 학습

- 중요 주파수 영역을 탐색하기 위해 어텐션 메커니즘을 사용
- **중요한 주파수** 일수록 높은 어텐션 스코어 $\alpha_{t,f}$ 값이 부여
- 중요 주파수 정보를 담은 새로운 주파수 히든 벡터 s_t 생성

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

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

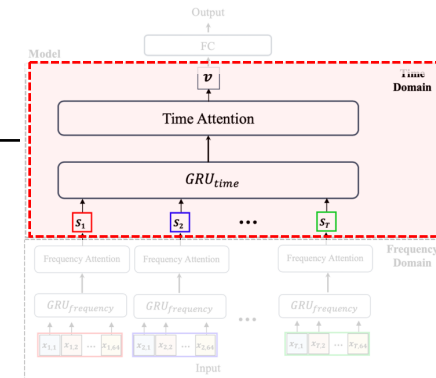
$$s_{t,f} = \sum_f \alpha_{t,f} h_{t,f}$$



04 제안 방법론

❖ 모델링 - ② 시간 구간 학습

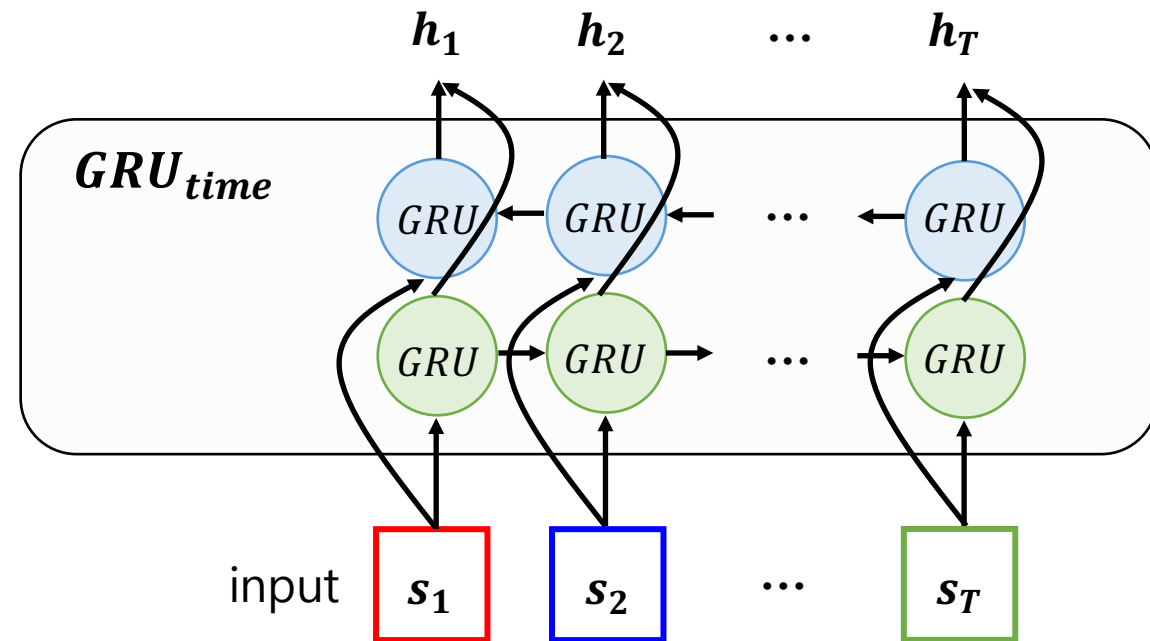
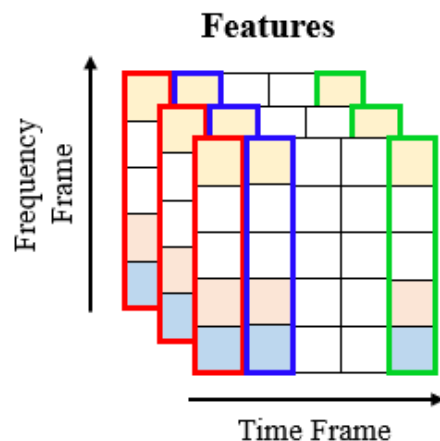
- 새로운 주파수 히든 벡터 s_t 를 **Bidirectional GRU**에 입력
- 정방향과 양방향 히든 벡터가 합쳐진 하나의 히든 벡터 출력 h_t



$$\vec{h}_t = GRU_{frequency}(s_t)$$

$$\overleftarrow{h}_t = GRU_{frequency}(s_t)$$

$$h_t = [\vec{h}_t, \overleftarrow{h}_t]$$



04 제안 방법론

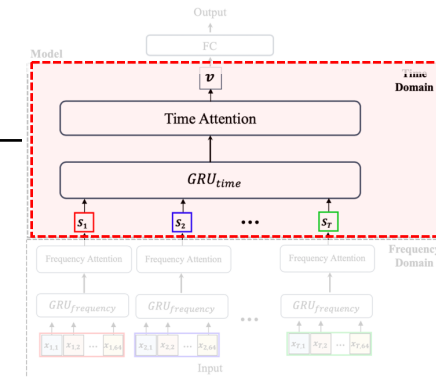
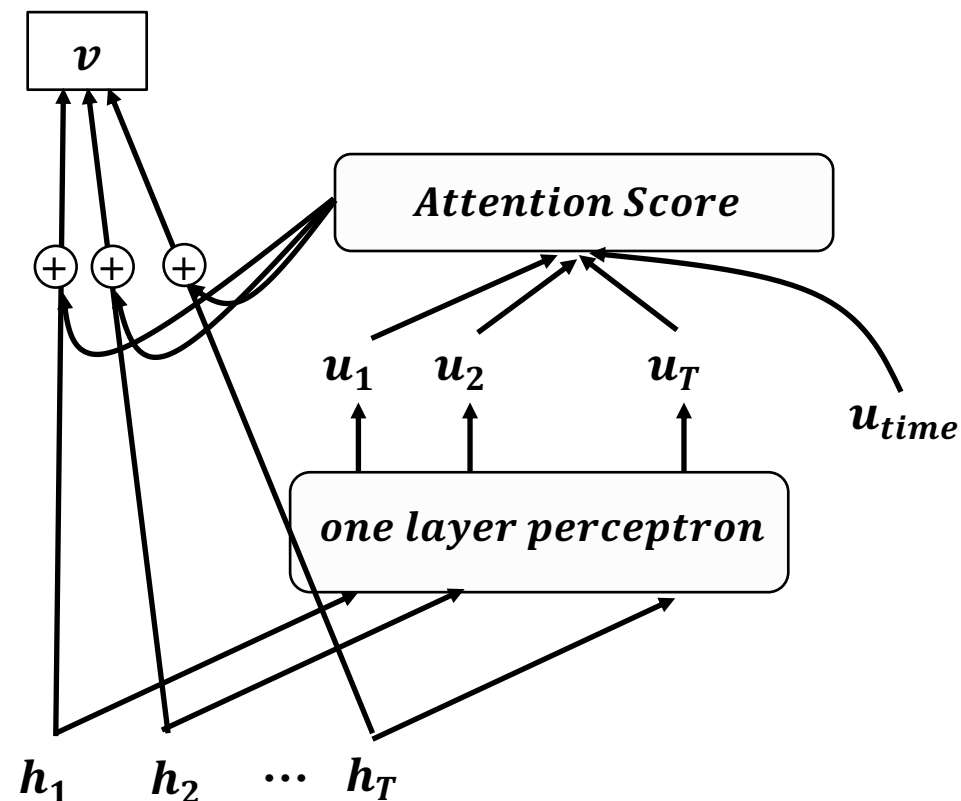
❖ 모델링 - ② 시간 구간 학습

- 중요 시간 영역을 탐색하기 위해 어텐션 메커니즘 사용
- **중요한 시간 일수록** 높은 어텐션 스코어 α_t 값이 부여
- 중요 주파수, 시간 정보를 담은 새로운 히든 벡터 v 생성

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

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

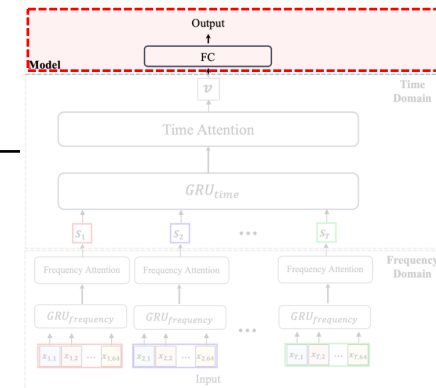
$$v = \sum_t \alpha_t h_t$$



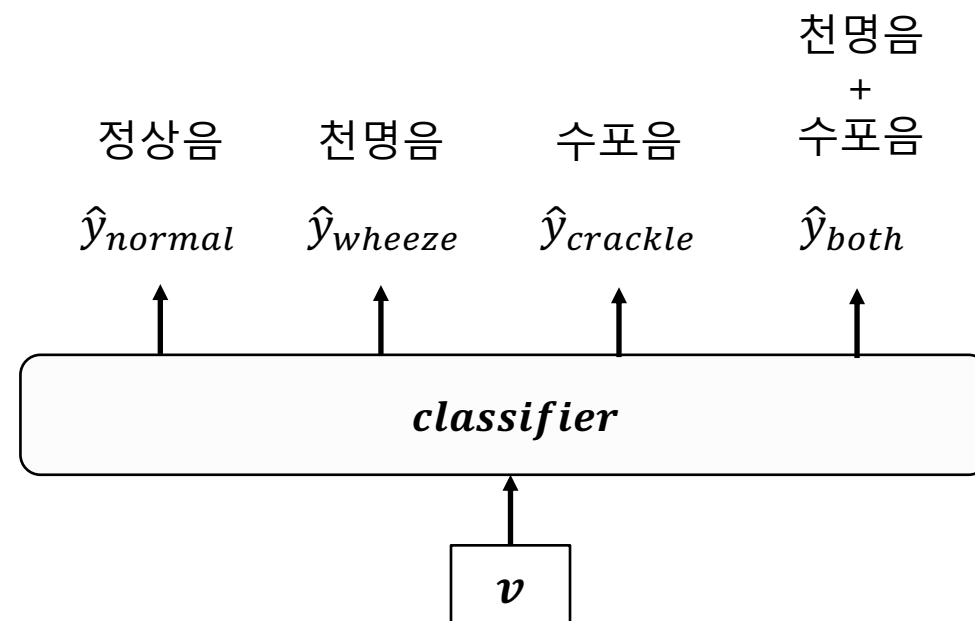
04 제안 방법론

❖ 모델링 - ③ 분류기

- 히든 벡터 히든 벡터 v 를 완전 연결층 레이어에 입력
- 출력된 값에 소프트 맥스 함수를 취해 각 클래스 확률 값 $\hat{y}_c$ 산출
- 가장 높은 확률 값으로 호흡음 클래스를 분류



$$\hat{y}_c = \text{softmax}(W_{\text{classifier}}v + b)$$



04 제안 방법론

❖ 하이퍼 파라미터 설정

- 주파수 영역 Bidirectional GRU 히든 차원 크기 : 50
- 시간 영역 Bidirectional GRU 히든 차원 크기 : 100
- 옵티마이저(Optimizer) : AdamW 알고리즘
- 학습률(Learning rate) : 0.001
- 손실 함수 : Cross Entropy

Cross Entropy Formulation

$$L(\hat{y}_c, class) = -\log\left(\frac{\exp(\hat{y}_{class})}{\sum_j \exp(\hat{y}_j)}\right) = -\hat{y}_{class} + \log\left(\sum_j \exp(\hat{y}_j)\right)$$

05 실험 결과

❖ 실험 설계

- Training / Testing set은 ICBHI 2017 Challenge에서 정의한 기준으로 분리
 - Training : 4,412 / Testing : 2,756
- 비교 대상 알고리즘
 - Support Vector Machine(SVM) / Random Forest(RF)
 - Residual network(RN)
 - ICBHI Challenge 2017 발표 방법론
 - 최근 제안된 딥러닝 기반 방법론
 - Ablation Study – Attention Mechanism

< Summary of the training, testing sets >

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

05 실험 결과

❖ 평가 지표

- ICBHI 2017 Challenge에서 정의한 평가 지표 사용

정상 호흡음 정확도

$$SP = \frac{N_n}{N}$$

이상 호흡음 정확도

$$SE = \frac{C_c + W_w + B_b}{C + W + B}$$

산술 평균

$$AS = \frac{SP + SE}{2}$$

조화 평균

$$HS = \frac{2 * SP * SE}{SP + SE}$$

< Determination rules >

(N: 정상음, W: 천명음, C: 수포음, W+C: 천명음 + 수포음)

		예측 클래스			
		N	W	C	W+C
실제 클래스	N	N_n	N_w	N_c	N_b
	W	W_n	W_w	W_c	W_b
	C	C_n	C_w	C_c	C_b
	W+C	B_n	B_w	B_c	B_b

05 실험 결과

정상 호흡음 정확도

$$SP = \frac{N_n}{N}$$

이상 호흡음 정확도

$$SE = \frac{C_c + W_w + B_b}{C + W + B}$$

산술 평균

$$AS = \frac{SP + SE}{2}$$

조화 평균

$$HS = \frac{2 * SP * SE}{SP + SE}$$

❖ 분류 성능 평가 결과

- 다른 방법론에 비해 이상 호흡음 정확도와 조화 평균이 높음
 - 이상 호흡음 정확도 : 34.75 %
 - 조화 평균 : 45.51 %

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

05 실험 결과

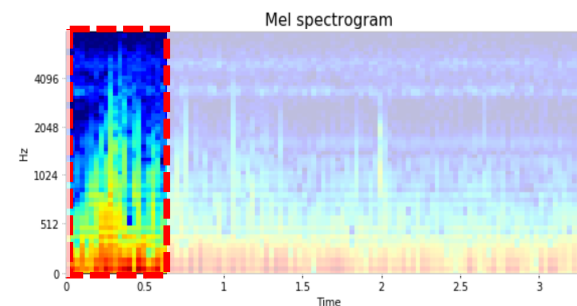
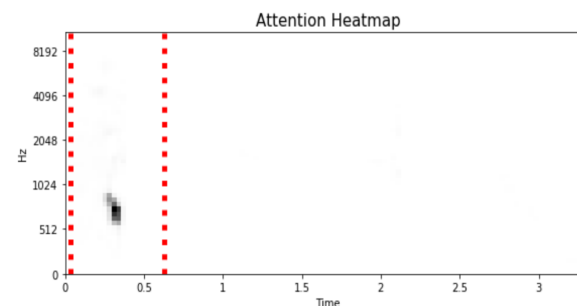
❖ 이상 호흡음 중요 정보 제공

- 시간, 주파수 영역 어텐션 스코어 연산으로 전체 영역에 대한 어텐션 스코어 $\alpha'_{t,f}$ 산출
 - 주파수 영역 어텐션 스코어 : $\alpha_{t,f}$ / 시간 영역 어텐션 스코어 : α_t
 - $\alpha'_{t,f} = \alpha_t * \alpha_{t,f} \quad \forall t \in [1, T]$
- 호흡기 질환은 소리 및 주파수 정보를 이용해 진단

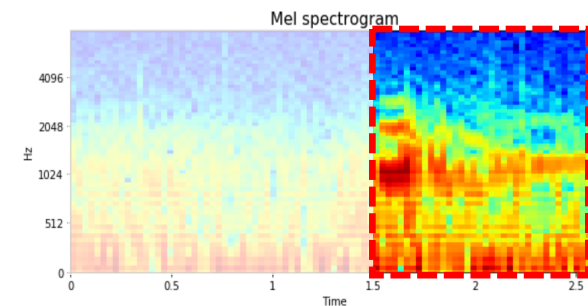
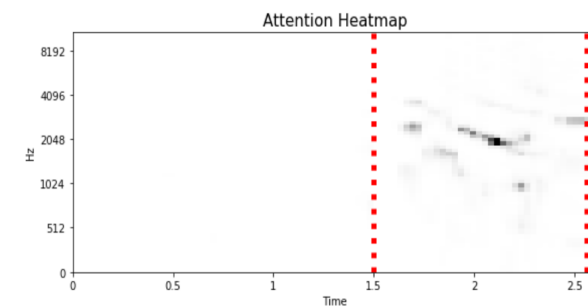
<예시>

만성폐쇄성질환 환자의 경우,

- 수포음(20ms 이하) : 호흡 주기 초반에 발생
- 천명음(100ms 이하) : 복조성(다양한 주파수) 형태를 띠



수포음



천명음

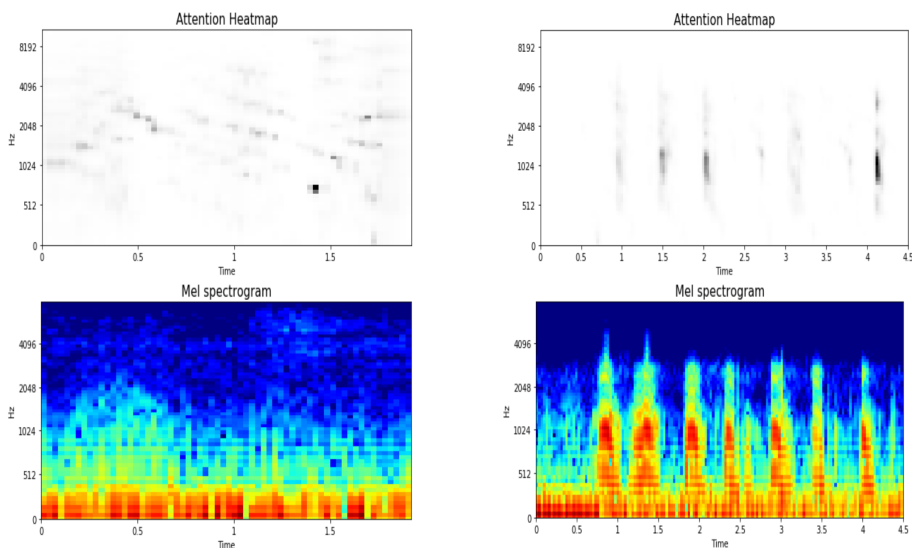
05 실험 결과

❖ 어텐션 스코어 시각화

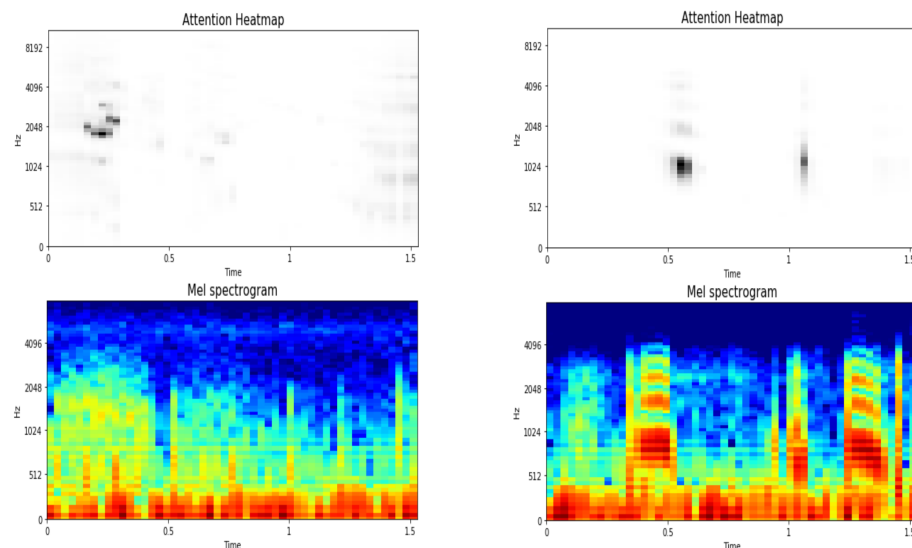
- 정상 호흡음의 경우, 어텐션 스코어가 특정 영역이 아닌 고르게 분포
- 노이즈와 같은 잡음에 의해 모델이 잘못 예측하는 경우가 발생

✓ 정상 호흡음 결과 예시

제대로 예측한 경우



잘못 예측한 경우



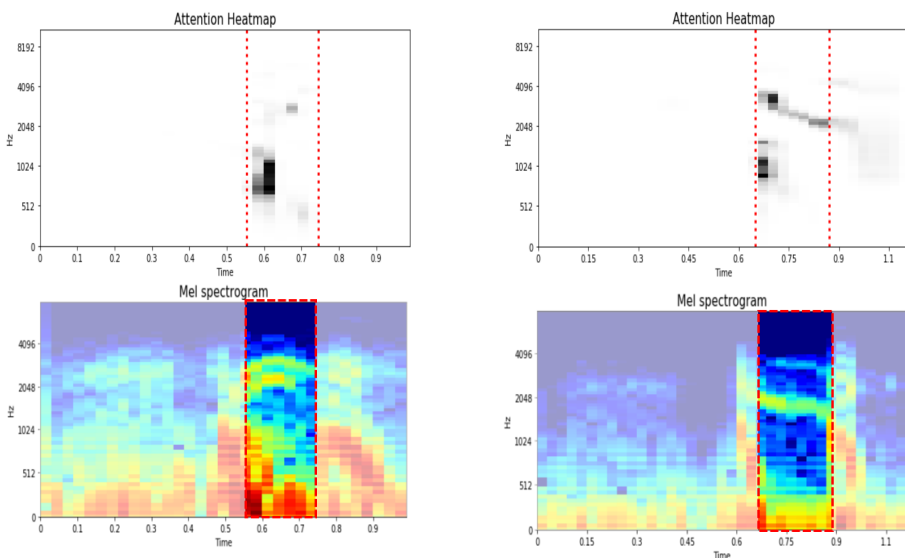
05 실험 결과

❖ 어텐션 스코어 시각화

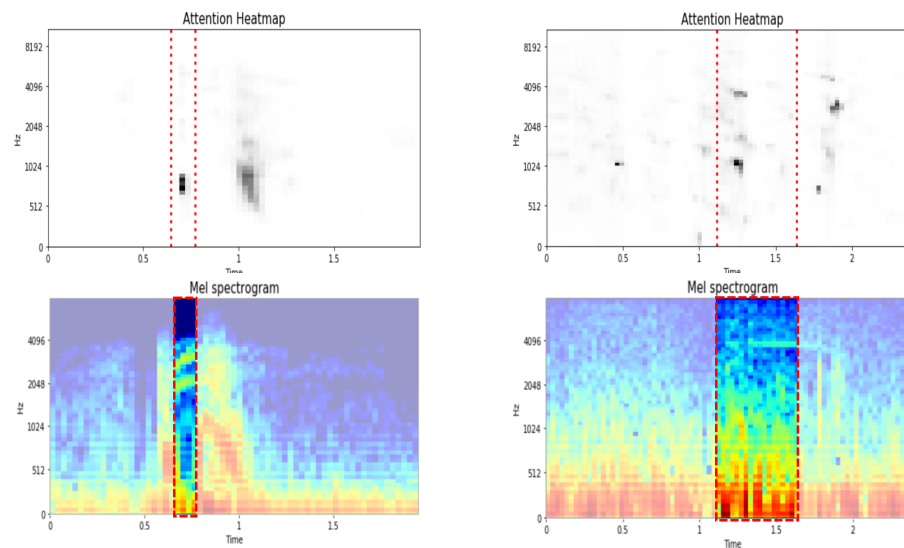
- 이상 호흡음의 경우, 이상 호흡음 발현 시간대에 어텐션 스코어가 높게 나타남
- 모델이 잘못 예측한 경우에도 어텐션 스코어가 적절하게 산출된 것을 확인할 수 있음

✓ 천명음 결과 예시

제대로 예측한 경우



잘못 예측한 경우



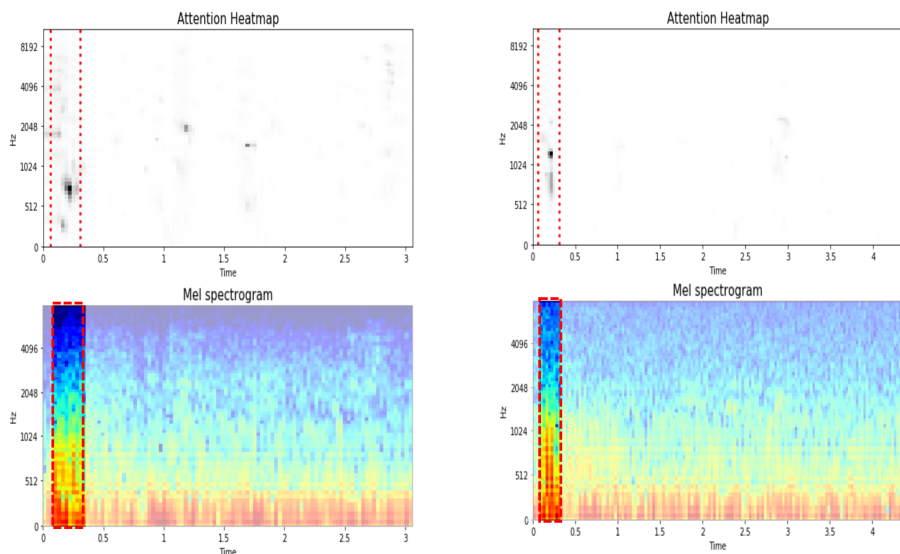
05 실험 결과

❖ 어텐션 스코어 시각화

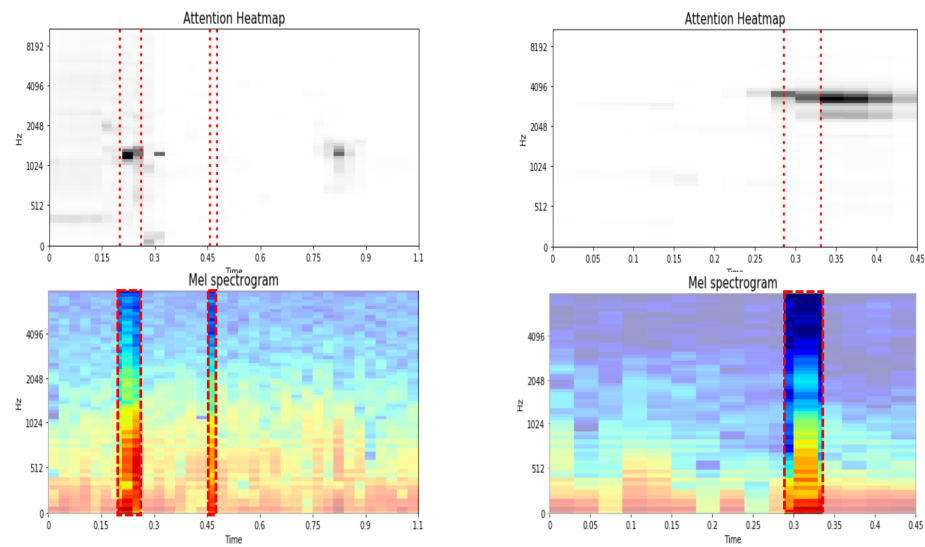
- 이상 호흡음의 경우, 이상 호흡음 발현 시간대에 어텐션 스코어가 높게 나타남
- 천명음에 비해 짧은 지속 시간을 보이는 수포음은 시간 영역 어텐션 스코어 형태가 짧게 나타남

✓ 수포음 결과 예시

제대로 예측한 경우



잘못 예측한 경우



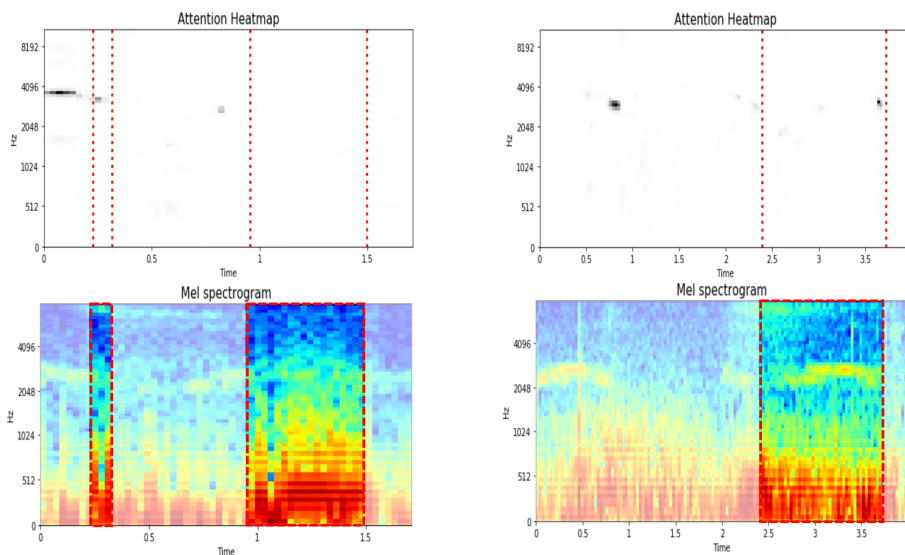
05 실험 결과

❖ 어텐션 스코어 시각화

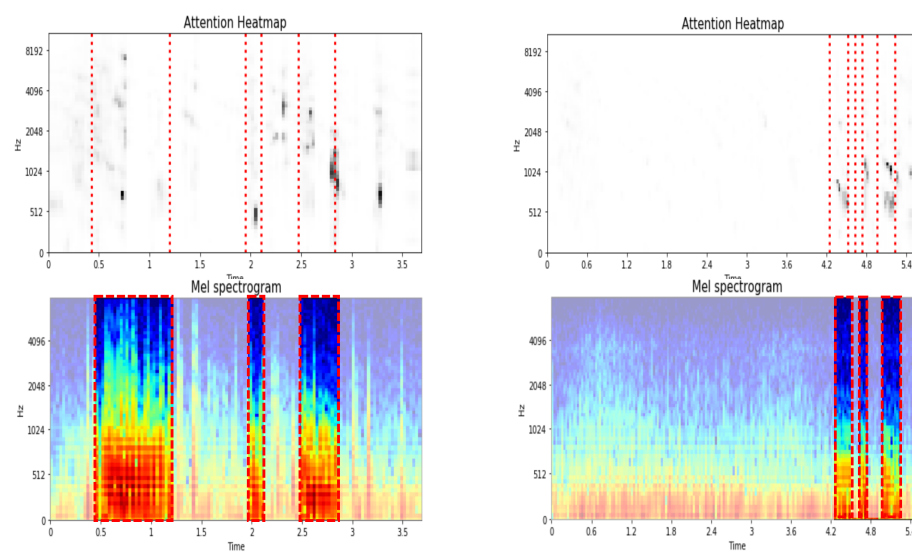
- 천명음과 수포음이 동시에 발생하는 호흡음에서도 적절하게 어텐션 스코어가 산출되는 것을 확인

✓ (천명음 + 수포음) 결과 예시

제대로 예측한 경우



잘못 예측한 경우



- 소리를 구성하는 시간과 주파수 영역을 계층적으로 학습하는 딥러닝 방법론 제안
 - 다른 방법론과 비교해 이상 호흡음에 대해 우수한 분류 성능을 확인
 - 어텐션 메커니즘을 통해 이상 호흡음의 중요 시간과 주파수 정보를 제공
 - 실제 의료 현장에서 실용적인 효과를 보일 수 있을 것으로 기대
- 소리가 발생하는 다양한 산업 현장에 적용할 수 있는 방법론이라는 점에서 큰 기여를 할 수 있을 것으로 기대
 - 제조 공정 과정에서 발생하는 소음을 이용한 불량품 감지
 - 불량품 발생 소음의 중요 시간과 주파수 영역 탐색을 통해 적절한 실험 설계 방안 제시
 - 사람 별 주요 주파수 형태를 탐색하여 음성 인식에 도움

참고 문헌

- [1] Breiman L. (2001). Machine Learning, 45(1), 5–32. Statistics Department, University of California, Berkeley, CA 94720. <https://doi.org/10.1023/A:1010933404324>.
- [2] Chambres, G., Hanna, P., & Desainte-Catherine, M. (2018). Automatic detection of patient with respiratory diseases using lung sound analysis. In Proceedings - International Workshop on Content-Based Multimedia Indexing. <https://doi.org/10.1109/CBML.2018.8516489>.
- [3] Cortes, C., & Vapnik, V. (1995). Support-Vector Cortes, C., & Vapnik, V. (1995). Support Vector Networks. Machine Learning, 20(3), 273–297. <http://doi.org/10.1023/A:1022627411411Networks>.
- [4] Faustino, P. S. (2019). Crackle and wheeze detection in lung sound signals using convolutional neural networks.
- [5] Jakovljević, N., & Lončar-Turukalo, T. (2018). Hidden Markov model based respiratory sound classification. In IFMBE Proceedings. <https://doi.org/10.1007/978-981-10-7419-67>.
- [6] Liu, C., Springer, D., Li, Q., Moody, B., Juan, R. A., Chorro, F. J., ... Clifford, G. D. (2016). An open access database for the evaluation of heart sound algorithms. Physiological Measurement. <https://doi.org/10.1088/0967-3334/37/12/2181>.
- [7] Ma, Y., Xu, X., Yu, Q., Zhang, Y., Li, Y., Zhao, J., & 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, Proceedings. <https://doi.org/10.1109/BIOCAS.2019.8919021>.
- [8] Minami, K., Lu, H., Kim, H., Mabu, S., Hirano, Y., & Kido, S. (2019, October). Automatic classification of large-scale respiratory sound dataset based on convolutional neural network. In 2019 19th International Conference on Control, Automation and Systems (ICCAS) (pp. 804-807). IEEE.
- [9] Palaniappan, R., & Sundaraj, K. (2013). Respiratory sound classification using cepstral features and support vector machine. 2013 IEEE Recent Advances in Intelligent Computational Systems, RAICS 2013. <https://doi.org/10.1109/RAICS.2013.6745460>.
- [10] Rocha, B. M., Filos, D., Mendes, L., Serbes, G., Ulukaya, S., Kahya, Y. P., ... De Carvalho, P. (2019). An open access database for the evaluation of respiratory sound classification algorithms. Physiological Measurement. <https://doi.org/10.1088/1361-6579/ab03ea>.
- [11] Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. Stata Journal. <https://doi.org/10.1177/1536867X20909688>.
- [12] Xie, S., Jin, F., Krishnan, S., & Sattar, F. (2012). Signal feature extraction by multi-scale PCA and its application to respiratory sound classification. Medical and Biological Engineering and Computing. <https://doi.org/10.1007/s11517-012-0903-y>.
- [13] Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & 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, NAACL HLT 2016 - Proceedings of the Conference. <https://doi.org/10.18653/v1/n16-1174>

감사합니다