Efficient Respiratory Disease Diagnosis Using Multi-Layered Audio Processing

Patrick Macasaet, Undergraduate

Department of Mechanical Engineering, Cullen College of Engineering

Problem Statement

- The continued rise of COVID-19 has exacerbated the impacts of respiratory diseases amongst older adults, developing children, and areas lacking proper healthcare infrastructure.
- Current methods of lung auscultation are primarily conducted with stethoscopes, and rely heavily upon physician arbitration, leaving the possibility for errors produced by stethoscope placement and audio clarity.
- The proposed methods utilize a novel spectrogram layering technique that is easy to process using lightweight algorithms designed for mobile use.

Brief Literature Review

- Wanasinghe et al. [1] uses a similar spectrographic layering technique, but excludes a COVID-19 class. Additionally, there is no colormapping applied to the final processed image and does not use gamma filter data.
- Patented device design by Samay Inc [2] implements a machine learning algorithm into a non-invasive wearable device, but is primarily tuned to detect COPD.

Datasets used

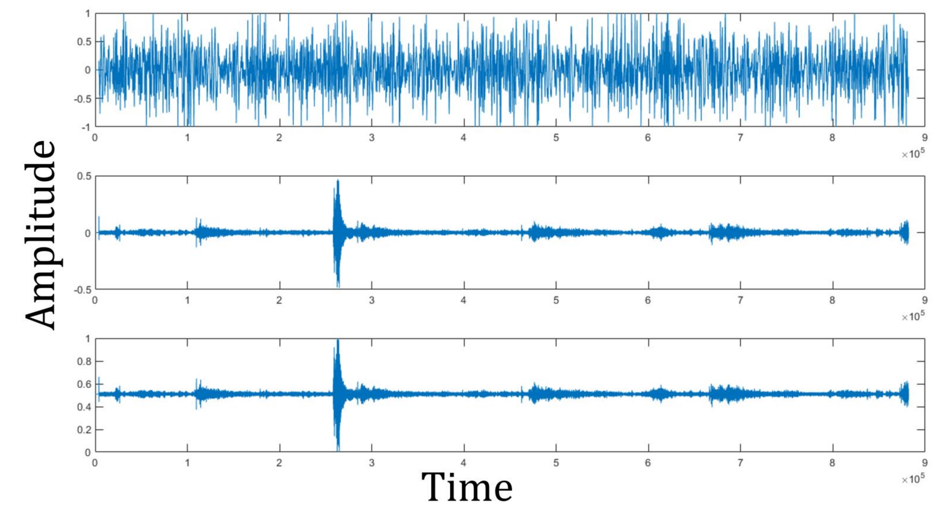
TEST/TRAINING DISTRIBUTION ACROSS ALL DISEASE CLASSES.

Category	Total Samples	Test/Train Split (80/20)
COPD	905	724/181
Pneumonia	435	348/87
COVID-19	655	524/131
Healthy	970	776/194

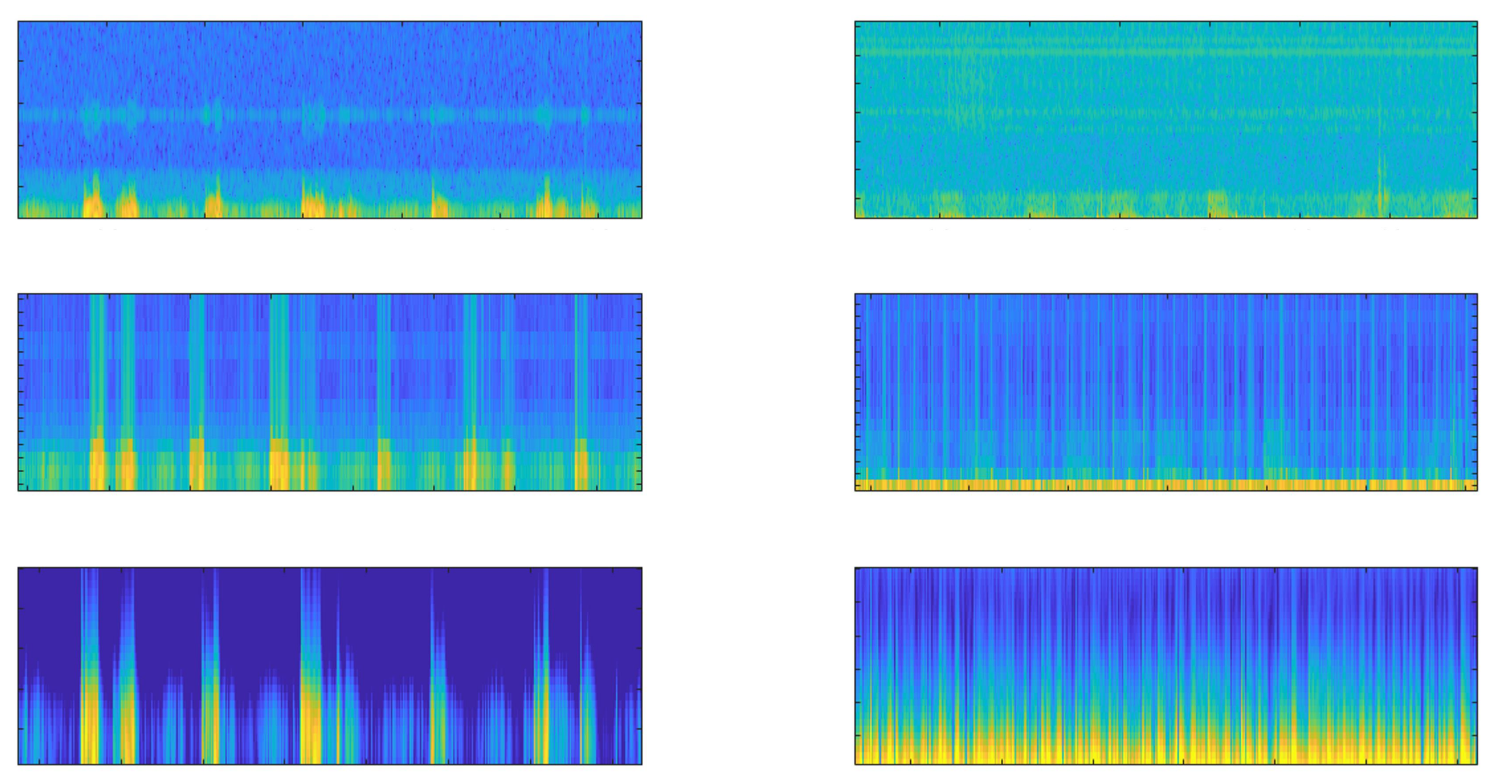
- The ICBHI database has over five hours of audio recordings from patients. Each audio file has about 20 seconds of recorded material from a range of electronic stethoscopes, with varying amounts of noise.
- The database created by Fraiwan et. al was used to augment the pneumonia class in this study. Recordings ranged from 5-30 seconds, and were recorded using a 3M™Littmann® Electronic Stethoscope 3200.
- **The open Coswara dataset was integral to making a novel contribution, since data includes user-submitted recordings using devices that range from stethoscopes to cellphone microphones. This is the only database that includes COVID-19 data.**
- For this project, only the breathing data pertaining to pneumonia, healthy, COPD, and COVID-19 to target diseases that have been at a risk of misdiagnosis since the introduction of COVID-19 [3][4].

Methods

- All audio was downsampled to 6000Hz
- A 6th order Butterworth bandpass filter with a low frequency cutoff of 50Hz and high cutoff of 2500Hz was applied.

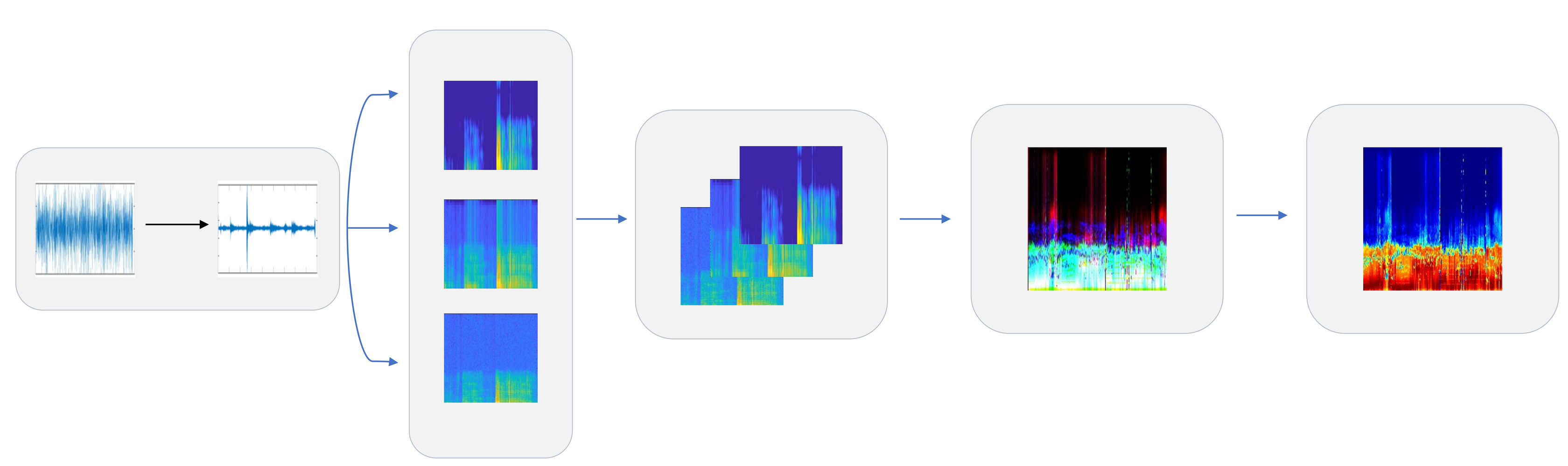


- The mel-spectrogram, gammatonegram, and spectrogram were used for visualizing and extracting the audio features associated with each class.
- **The novel combination of the mel-spectrogram and gammatonegram data reflects the features on a scale more relevant to human auditory perception.**

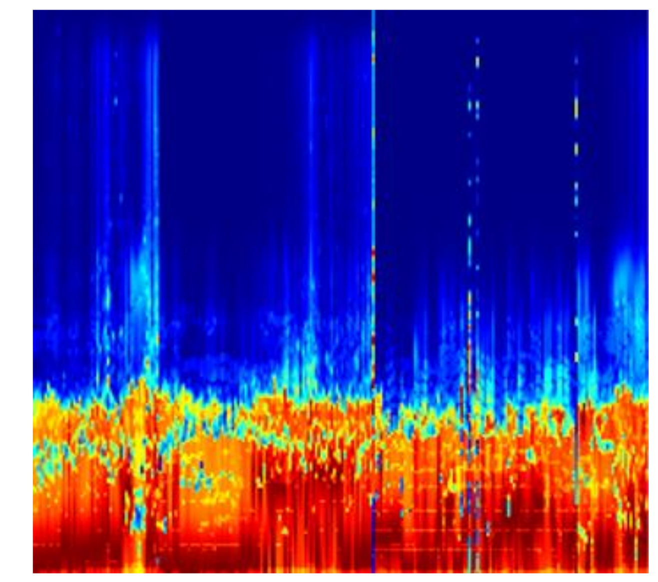


COPD **Healthy**

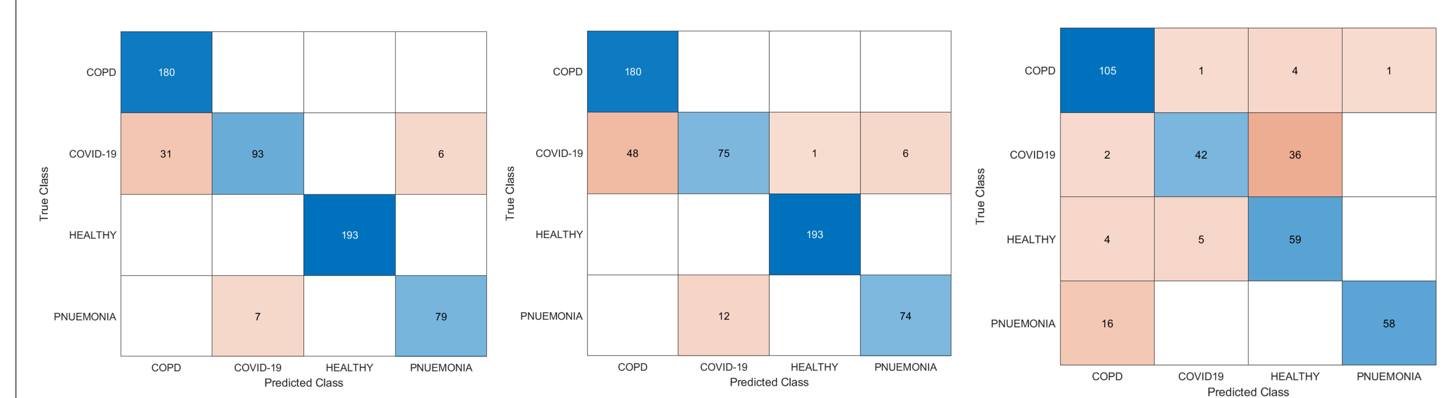
- Each spectral representation of the normalized audio was layered into a 3-channel RGB image, then **an additional colormap was applied to further develop extracted features.**



Results, Analysis and Discussion



- The resulting images were input through ResNet50, MobileNetV3, and Squeezenet, **prioritizing lightweight CNN architectures designed for mobile implementation.**
- Training lasted for ten epochs to reduce overfitting and training time.



- MobileNet had the strongest performance, followed by ResNet and Squeezenet; training time was under ten minutes total for all.

TABLE II

ACCURACY, F1-SCORE, AND PRECISION METRICS ACROSS CLASSES.

	Accuracy	F1-score	Precision
MobileNetv3	0.925	0.891	0.853
ResNet50	0.886	0.843	0.789
Squeezenet	0.792	0.753	0.827

- Generally, the processed audio dataset performed well with lightweight CNN models, achieving **high accuracy with minimal training epochs.**
- Desirable traits of this training pipeline also include low storage space **(each trained CNN model is under 150MB, full image dataset is under 50MB).**

Conclusions

- **Efficiency points towards potential integration into low-cost, resource-efficient point-of-care diagnostic devices not yet seen on the market.**
- **The training of the CNN does not require the presence of the original audio file for training, cutting down the size of the training set from multiple gigabytes to potentially a few hundred kilobytes.**
- **Novel inclusion of the mel-spectrogram and gammatonegram provides a better frequency representation that reflects human auditory perception.**
- **Additional colormapping added to layered spectrogram representation coupled with the mel-gammatone combination increased accuracy of pre-loaded MobileNetV3 module near levels of current studies [1].**

References (brief)

1. T. Wanasinghe, S. Bandara, S. Madusanka, D. Meedeniya, M. Bandara and I. D. L. T. Díez, "Lung Sound Classification With Multi-Feature Integration Utilizing Lightweight CNN Model," in IEEE Access, vol. 12, pp. 21262-21276, 2024, doi: 10.1109/ACCESS.2024.3361943.
2. Artunduaga, M. A. (2023, May 16). Artunduaga, M. (2023, May 16). Systems, devices, and methods for performing active auscultation and detecting sonic energy measurements.
3. Hadavand, F., Shoaei, S. D., \& Kharazmi, A. B. (2023). Misdiagnosed Pneumocystis Pneumonia as COVID-19: A Case Report. Tanaffos, 22(2), 272–275.
4. ELENA AVALOS PEREZ-URRIA, María Rodrigo-García, Adrián Peláez, Laura Castellanos López, Alberto Martínez De Lara, Elena García Castillo, Tamara Alonso Pérez, Celeste Marcos, Rosa María Girón Moreno, Claudia Valenzuela, Julio Ancochea, Joan B Soriano European Respiratory Journal Sep 2023, 62 (suppl 67) PA2400; DOI: 10.1183/13993003.congress-2023.PA2400

Acknowledgements

The financial support from Midstream Integrity Services (MIS) and technical support from Smart Materials & Structures Lab (SMSL) and Artificial Intelligence Lab for Monitoring & Inspection (AILMI) at UH.