Detecting Disease from Respiratory Audio

Mithun Subhash, Aryan Mittal, Akshay Raj, Arush Verma, Prajwal Mohan Kumar

Introduction

With the rate of respiratory ailments on the rise, the stress imposed on global health becomes more prevalent. Recent developments in medical technology have yielded datasets of recorded lung sounds, opening up new possibilities for automated lung sound examination.

Motivated by our desire to improve early detection of serious respiratory disease, we look to differentiate between 6 kinds of respiratory conditions. These include:

- COPD
- URTI
- Bronchiectasis
- Pneumonia
- Bronchiolitis
- Healthy

We hope to utilize this in a web application for easy at home screening.

Dataset

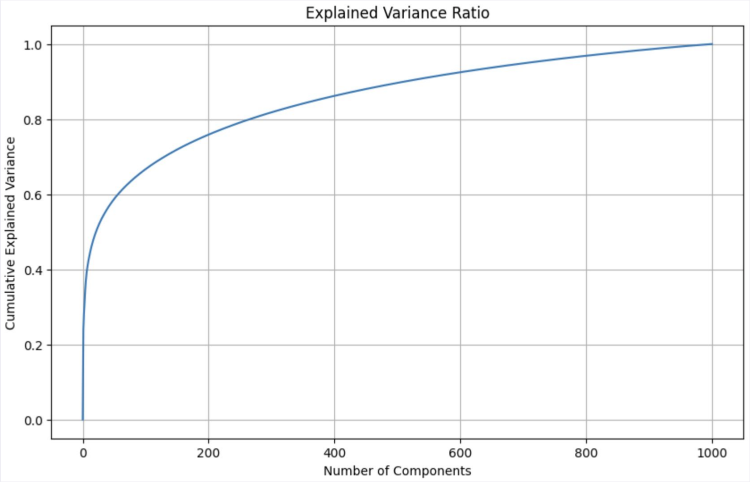
- 920 wav files
  - 10 to 90 seconds
  - Nearly 7000 breathing cycles
- 920 corresponding txt files
  - Have patient data
  - Have associated information about the disease

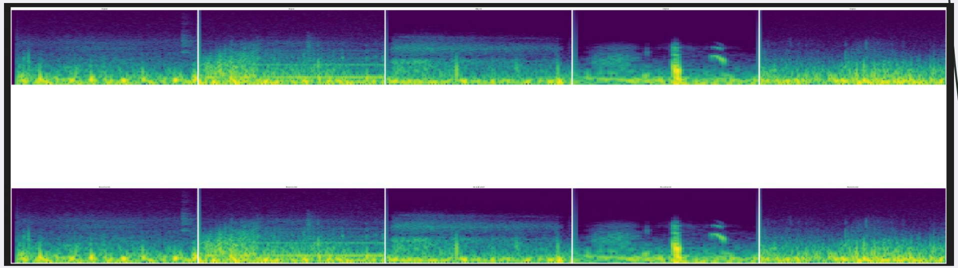
```
The demographic info file has 6 columns:
 - Patient number
 - Age
 - Sex
 - Adult BMI (kg/m2)
 - Child Weight (kg)
 - Child Height (cm)


Each audio file name is divided into 5 elements, separated with underscores (_).

1. Patient number (101,102,...,226)
2. Recording index
3. Chest location
      a. Trachea (Tc)
      b. Anterior left (Al)
      c. Anterior right (Ar)
      d. Posterior left (Pl)
      e. Posterior right (Pr)
      f. Lateral left (Ll)
      g. Lateral right (Lr)
4. Acquisition mode
     a. sequential/single channel (sc),
     b. simultaneous/multichannel (mc)
5. Recording equipment
     a. AKG C417L Microphone (AKGC417L),
     b. 3M Littmann Classic II SE Stethoscope (LittC2SE),
     c. 3M Litmmann 3200 Electronic Stethoscope (Litt3200),
     d.  WelchAllyn Meditron Master Elite Electronic Stethoscope (Meditron)

The annotation text files have four columns:
- Beginning of respiratory cycle(s)
- End of respiratory cycle(s)
- Presence/absence of crackles (presence=1, absence=0)
- Presence/absence of wheezes (presence=1, absence=0)

The abbreviations used in the diagnosis file are:
- COPD: Chronic Obstructive Pulmonary Disease
- LRTI: Lower Respiratory Tract Infection
- URTI: Upper Respiratory Tract Infection
```

Preprocessing

- Created audio clips for each breathing cycle which could be analyzed individually.
- Generated mel spectrograms, where x axis is time and y-axis is frequency.
- Joined demographic data with cycle data for easy lookup in dataframe.
- Used butter band-pass filter that removed background noise and heartbeat on clips.

We also used PCA on the mel spectrograms to reduce the amount of features. We retained 99% of original variance as audio classification relies on small difference. At 927 components we reached 99% retained variance. To the right, you can see original spectrograms on the top and reconstructed ones on the bottom.





Methods

01 GMM

- Helps find patterns in data through clustering.
- Combination of gaussian models is good at finding clusters in demographic data.

02 DBSCAN

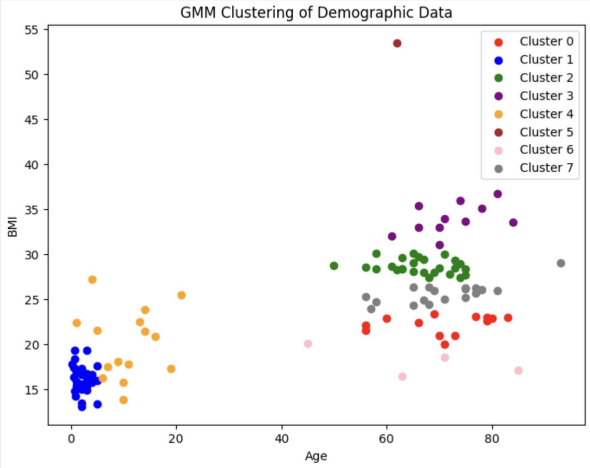
- Good for detecting arbitrary clusters
- Can find irregular patterns

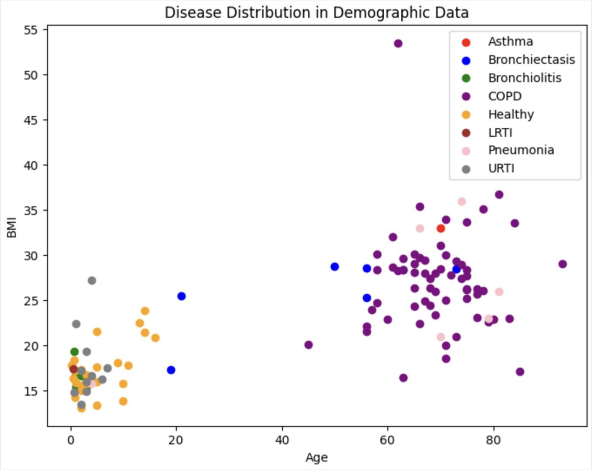
03 CNN

- Good for recognizing complex features
- Convolutional layers provide flexibility in learning
- Good for audio classification

GMM Results

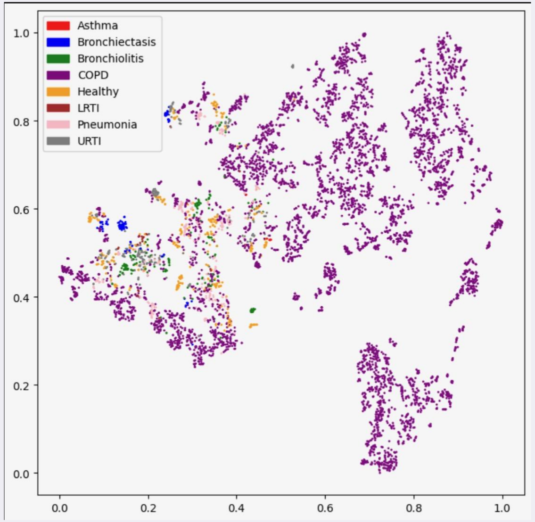
- We ran GMM on the most relevant demographic features (age and bmi).
- We used both an internal and an external measure to measure performance
- Internal – Silhouette Coefficient – We got a value of 0.376, which is less than the desired value of 0.5. THis means our clusters are not well defined.
- External – Completeness Score – We got a value of 0.388, which is not good. This metric reflects how good clusterings are with respect to true labels.
- GMM appears to not produce valuable results in this use case.

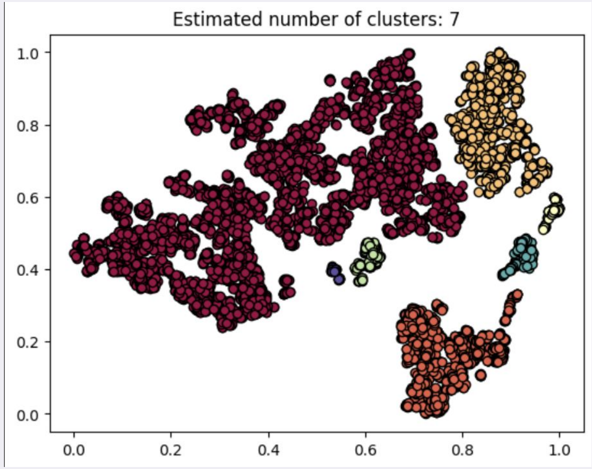




DBSCAN Results

- Extracted 13 MFCCs from each audio clip
- Dimension reduced 13 to 2 for plotting using t-SNE
- Clustered the t-SNE datapoints using DBSCAN
- Best results obtained with min_pts=5, epsilon=.035
    - silhouette coefficient = .074
    - completeness score = .031
- DBSCAN seems to produce worse results than GMM with respect to disease prediction





Unsupervised Comparison

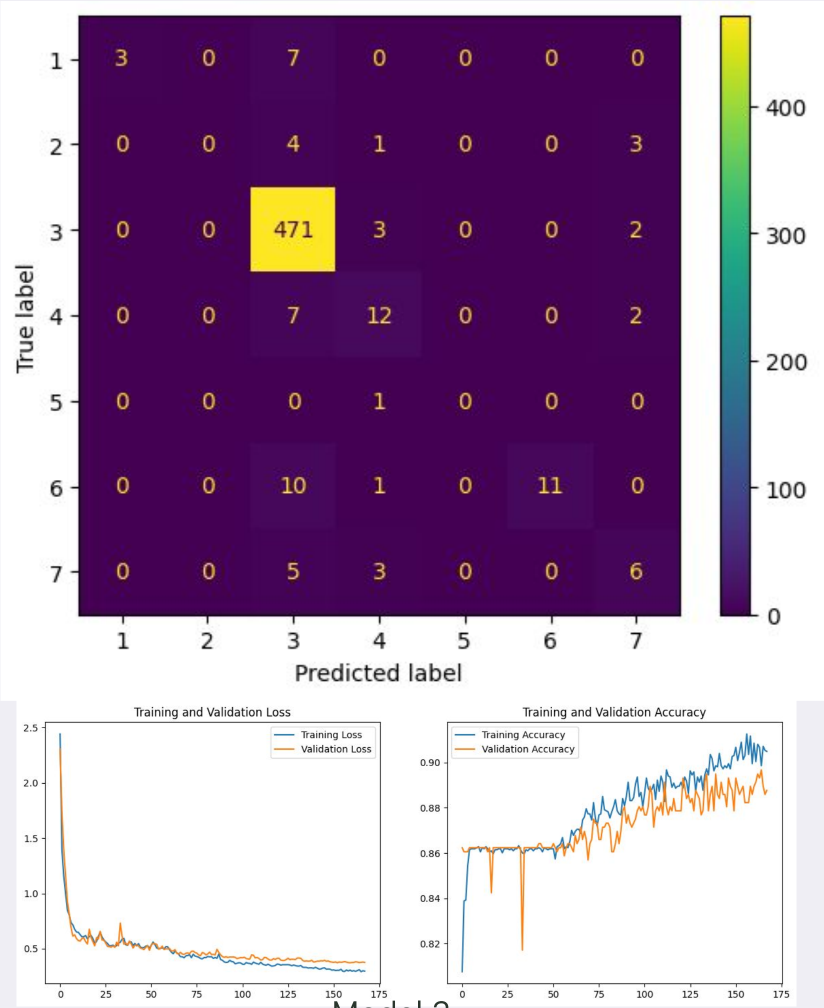
- Completeness scores far higher (10x) when using GMM on the demographic data than for DBSCAN on the audio files
- Demographic data is a better predictor of disease than the audio files

CNN Results

- We ran CNN to classify the respiratory audio diseases

[Bronchiectasis, Bronchiolitis, COPD, Healthy, LRTI, Pneumonia, URTI]

- Model 1 Accuracy: 88.77%
  - Layers: 1D Convolutions, 1D Max-Pooling, Dropout/Dense/Flatten
- Model 2 Accuracy: 90.6%
  - Changes: Residual Block, Batch Normalization, Regularization, Dynamic Learning Rate
- Model 3 Accuracy: 91.3%
  - Changes: Increase dropout, , increase regularization parameter
- Conclusion
  - Model 3 had a 2.85% improvement from the original model
  - CNN is a viable method to predict respiratory audio diseases



Model 3

Next Steps

- Short term
  - Try out different models such as random forest to model the mix of categorical and continuous data
  - Try advanced architectures, such as Transformer-based (AST, ViT) or Mamba-based (Vision Mamba)
  - Develop better architecture for CNN

- Long Term
  - FInd or create another dataset that we can apply our models to to see if this can be applied to other datasets
  - Examine for other diseases

References

[1] S. Bae *et al.*, “Patch-Mix Contrastive Learning with Audio Spectrogram Transformer on Respiratory Sound Classification,” in *INTERSPEECH 2023*, ISCA, Aug. 2023, pp. 5436–5440. doi: 10.21437/Interspeech.2023-1426.

[2] S. Gairola, F. Tom, N. Kwatra, and M. Jain, “RespireNet: A Deep Neural Network for Accurately Detecting Abnormal Lung Sounds in Limited Data Setting,” in 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Nov. 2021, pp. 527–530. doi: 10.1109/EMBC46164.2021.9630091.

[3] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, “Class-Balanced Loss Based on Effective Number of Samples,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, pp. 9260–9269. doi: 10.1109/CVPR.2019.00949.

[4] “Respiratory Sound Database.” Accessed: Feb. 22, 2024. [Online]. Available: https://www.kaggle.com/datasets/vbookshelf/respiratory-sound-database

[5] B. M. Rocha et al., “A Respiratory Sound Database for the Development of Automated Classification,” in Precision Medicine Powered by pHealth and Connected Health, vol. 66, N. Maglaveras, I. Chouvarda, and P. De Carvalho, Eds., Singapore: Springer Singapore, 2018, pp. 33–37. doi: 10.1007/978-981-10-7419-6_6.

[6] A. H. Sabry, O. I. Dallal Bashi, N. H. Nik Ali, and Y. Mahmood Al Kubaisi, “Lung disease recognition methods using audio-based analysis with machine learning,” Heliyon, vol. 10, no. 4, p. e26218, Feb. 2024, doi: 10.1016/j.heliyon.2024.e26218.

[7] Y. Gong, Y.-A. Chung, and J. Glass, “AST: Audio Spectrogram Transformer,” in Interspeech 2021, ISCA, Aug. 2021, pp. 571–575. doi: 10.21437/Interspeech.2021-698.

[8] L. Fedden, “Comparative Audio Analysis With Wavenet, MFCCs, UMAP, t-SNE and PCA,” Medium, Nov. 21, 2017. https://medium.com/@LeonFedden/comparative-audio-analysis-with-wavenet-mfccs-umap-t-sne-and-pca-cb8237bfce2f (accessed Apr. 24, 2024).

[9] P. Zhang, A. Swaminathan, and A. Uddin, “Pulmonary disease detection and classification in patient respiratory audio files using long short-term memory neural networks,” Frontiers in Medicine, vol. 10, Nov. 2023, doi: https://doi.org/10.3389/fmed.2023.1269784

[10] L. Com and G. Hinton, “Visualizing Data using t-SNE Laurens van der Maaten,” Journal of Machine Learning Research, vol. 9, pp. 2579–2605, 2008, Available: https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf