Classification of Respiratory Sounds

Vinita Shivakumar
Stanford University
School of Medicine
vinshiv@stanford.edu

Abstract

Respiratory diseases are a leading cause of death in the world and accurate lung auscultation is extremely important for the diagnosis and evaluation of disease. However, this method is vulnerable to physician and instrument limitations and there is strong interest in automation of lung sound analysis. In this paper, I explore the use of previously described CNNs to classify two adventitious respiratory sounds – wheezes and crackles – using a publicly available respiratory sounds database that has previously been used for non-machine learning techniques. My results demonstrate that CNNs can achieve much higher accuracy with fewer pre-processing steps and that CNNs are most effective when trained separately for different sounds.

1. Introduction

The lungs are one of the most vital organs in our body, yet one that we often take for granted. Respiratory diseases such as chronic obstructive pulmonary disease (COPD), asthma, acute lower respiratory tract infections, tuberculosis, and lung cancer are amongst the leading causes of death and disability worldwide [1]. Lung auscultation is a cheap, non-invasive, safe, and easy to perform diagnostic technique that is still a critical part of the physical exam today. However, it is still extremely vulnerable to false interpretations based on physician and instrument limitations [2]. There is significant clinical interest in computerized analysis of respiratory sounds but the use of machine learning algorithms in this field is still preliminary [3, 4, 5, 6].

Lung sounds being non-stationary and non-linear signals makes them difficult to analyze and distinguish and use of electronic stethoscope has made automated analyses possible. Past research have found CNNs to be an effective means of classifying abnormal breath sounds and [7].

Recently, the largest publicly available respiratory sound database was compiled to encourage the development of algorithms that can identify common adventitious breath sounds (wheezes and crackles) from clinical and non-clinical settings [8]. Crackles are discontinuous sounds, typically less than 20 ms, associated with lung fibrosis (fine crackles) or chronic airway obstruction (coarse crackles) [9,10]. Wheezes are high pitched sounds that usually last more than 100 ms that indicate obstructive airway conditions such as asthma and COPD [9, 10]. A total of 6898 respiratory cycles (a single inhalation and exhalation) were recorded and annotated by experts as wheezes, crackles, both, or no adventitious sounds (i.e. normal). Out of a total of 11 entries, the highest accuracy achieved was 50% using a Support Vector Machine (SVM) multi-class classifier [11] and there are currently no publications that use machine learning for this dataset. The goal of this project is to improve the accuracy of classification using convolutional neural networks (CNNs).

2. Related work

Recently, research on the use of CNNs in lung sound analysis has begun to be published [4, 5, 6, 12]. In 2017, one paper concluded that " spectrogram image classification with CNN algorithm works as well as the SVM algorithm, and given the large amount of data, CNN and SVM machine learning algorithms can accurately classify and pre-diagnose respiratory audio" [4]. However, this paper used the same CNN for 13 different respiratory sound types, all of which have different frequency and duration profiles. The only other publication of CNN classification of lung sounds in 2018 also created a single model to classify 7 different sound types [5]. Their conclusion was that "CNN outperformed the handcrafted feature based classifiers" namely SVM, k-nearest neighbor and Gaussian mixture models [5].

I utilize the CNNs proposed in these papers but create different models for each sound type.

3. Dataset

For this project, I use the Respiratory Sound Database from the International Conference on Biomedical and

Health Informatics 2017 Challenge. This is the first publicly available respiratory sounds dataset [10]. It contains 920 audio samples collected independently from two centers in Portugal and Greece. The files contain a total of 5.5 hours of recordings with 6898 respiratory cycles from 126 subjects.

The cycles were annotated by respiratory experts as including crackles, wheezes, a combination of them, or no adventitious respiratory sounds leading to 1864 crackles, 886 wheezes and 506 with both. The recordings were collected using heterogeneous equipment and their duration ranged from 10s to 90s. The chest locations from which the recordings were acquired is also provided. Noise levels in some respiration cycles is high, which simulate real life conditions. patients. Each annotation file contains four columns: the beginning and end of each respiratory cycle in the recording, presence of crackle or not and presence of wheeze or not.

4. Method

To summarize, most of the effort for this project went into pre-processing the audio files and developing the model.

Pre-processing

- Slice .wav files into individual respiratory cycles to correspond to annotations (separate wheeze and crackle output set)
- Resample to 4000 Hz to set up coherent feature set
- Apply 12th-order Butterworth bandpass filter with 120, 1800 Hz cut off frequencies to minimize noise effects
- Convert to mel- for input to CNN: created 128x16 spectrograms where 128 refers to number of mel windows and 16 refers to number of frames. I chose 128 as this is a standard number and 16 as I wanted to start with a small number of frames and then tune as a hyperparameter.

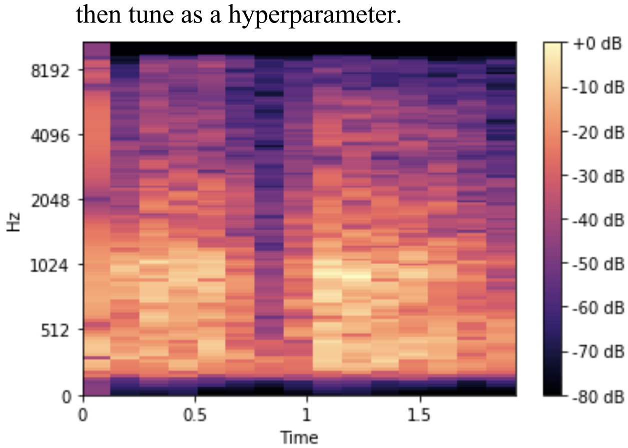

Fig 1: Example of mel-spectrogram of spliced audio file

Neural network: I modified an existing CNNs to create the base model for my dataset.

- Spectrogram dimensions 128x16x1 [4]
- 1st convolutional layer with filter size 3x3 and 32 feature maps
- Max pooling 2x2
- 2nd convolutional layer with filter size 3x3 and 32 feature maps. o Max pooling 2x2
- Flatten: Fully connected layer
- Output: single neuron for presence of specific sound (wheeze or crackle)

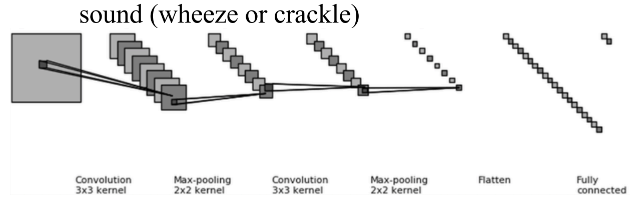

Fig 2: Picture of original CNN from 2017 paper [4]

For loss function, I used sigmoid cross entropy loss:

$$CE = -\sum_{i=1}^{C'=2} t_i log(s_i) = -t_1 log(s_1) - (1-t_1) log(1-s_1)$$

Furthermore I used an Adam optimizer with learning rate 0.009 and batch size of 64. For the first model, I used both wheezes and crackles simultaneously for 10 epochs. I then split the dataset and ran the model on wheezes and crackles separately again for 10 epochs. I used both a 90-10 and 80-20 train-test split – the results for both were the same and therefore only 90-10 results are shown here.

5. Results and Analysis

The first model had a train accuracy of 48.69% and a test accuracy of 50%. This was already performing at a higher level than the SVM initially used on the dataset.

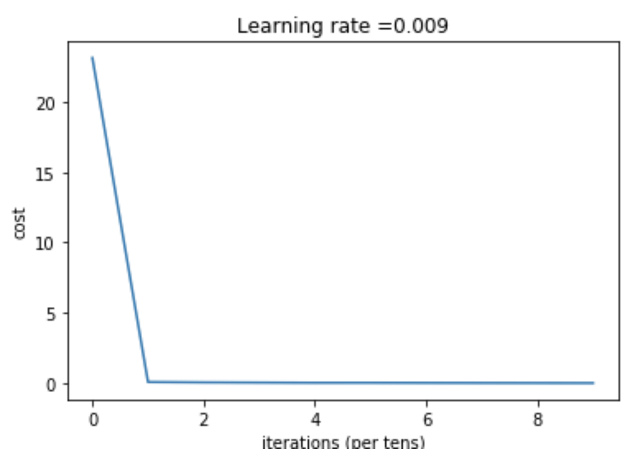

Fig 3: Cost function of model 1

The second model showed a train and test accuracy of 100%. This was an unexpected result and requires further investigation.

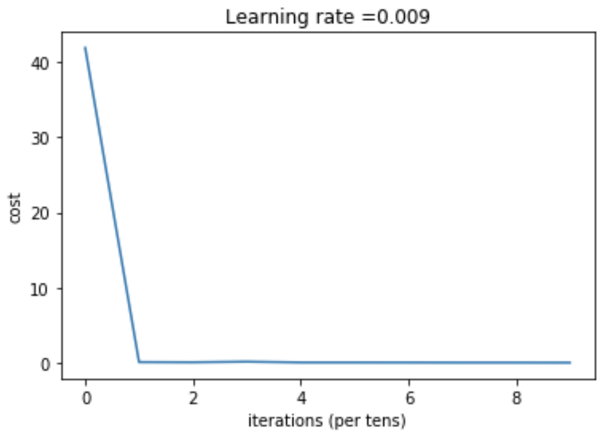

Fig 4: Cost function of model 2

This result clearly demonstrates that splitting up models for each sound type is very beneficial.

6. Conclusion and Future Work

This result clearly demonstrates that splitting the sounds up into different models is very beneficial. Future work includes testing the model with more data which presumably would reduce the accuracy from 100%. That would also allow hyperparameter tuning such as number of frames for the spectrogram. Further, I would consider training a deeper network similarly to that described in the 2018 paper [5]. Lastly, if no other data was available, I would consider data augmentation by time warp. This is one of four speech data augmentation methods – the only one of which seems to be appropriate for my data. However, this is reported to be the most computationally expensive and least effective method so this would be a last resort if I am unable to find more comparable data.

References

[1] Forum of International Respiratory Societies. The Global Impact of Respiratory Disease – Second Edition. Sheffield, European Respiratory Society, 2017.

[2] H. Pasterkamp, S.S. Kraman, G. Wodicika. "Respiratory sounds," *American Journal of Respiratory and Critical Care Medicine*, vol. 156, pp. 974-987, 1997.

[3] R. Palaniappan, K. Sundaraj, N.U. Ahamed. "Machine learning in lung sound analysis: A systematic review," *Biocybernetics and Biomedical Engineering*, vol. 33, no. 3, pp. 129-135, 2013.

[4] M. Aykanat, O. Kilic, B. Kurt, S. Saryal. "Classification of lung sounds using convolutional neural networks," *Eurasip Journal on Image and Video Processing*, vol. 65, 2017.

[5] D. Bardou, K. Zhang, S.M. Ahmad. "Lung sounds classification using convolutional neural networks," *Artificial Intelligence in Medicine*, vol. 88, pp. 58-69, 2018.

[6] C. Jacome, J. Ravn, E. Holsbo, J.C. Aviles-Solis, H. Melbye, L.A. Bongo. "Convolutional Neural Networks for Breathing Phase Detection in Lung Sounds," *Sensors (Basel)*, vol. 19, no. 8, pp. 1798, 2019.

[7] M. Gronnesby, J.C. Aviles-Solis, E. Holsbo, H. Melbye, L.A. Bongo. "Feature Extraction for Machine Learning Based Crackle Detection in Lung Sounds from a Health Survey," 2017.

[8] B.M. Rocha, D. Filos, L. Mendes, I. Vogiatzis, E. Perantoni, E. Kaimakamis, et al. "A Respiratory Sound Database for the Development of Automated Classification," *Precision Medicine Powered by Health and Connected Health*, pp. 33–37, 2018.

[9] M. Sarkar, I. Madabhavi, N. Niranjan, M. Dogra. "Auscultation of the respiratory system," *Annals of Thoracic Medicine*, vol. 10, no. 3, pp. 158-168, 2015.

[10] B.M. Rocha, D. Filos, L. Mendes, G. Serbes, S. Ulukaya, Y.P. Kahya, et al. "An open access database for the evaluation of respiratory sound classification algorithms," *Physiological Measurement*, vol. 40, no. 3, 2019.

[11] G. Serbes, S. Ulukaya, Y.P. Kahya. "An Automated Lung Sound Preprocessing and Classification System Based On Spectral Analysis Methods," *Precision Medicine Powered by Health and Connected Health*, vol. 66, pp. 45-49, 2018.