WheezeD: Respiration Phase Based Wheeze Detection Using Acoustic Data From Pulmonary Patients Under Attack

Soujanya Chatterjee*
University of Memphis
schttrj1@memphis.edu

Md. Mahbubur Rahman
Ebrahim Nemati
Jilong Kuang
Samsung Research America
{m.rahman2,e.nemati,jilong.kuang}@samsung.com

ABSTRACT

Wheezing is one of the most prominent symptoms for pulmonary attack. Hence, wheezing detection has attracted a lot of attention in recent years. However, there is a dearth of a reliable method that can automatically detect wheezing events during each respiration phase in presence of several concurrent sounds such as cough, throat clearing, and nasal breathing. In this paper, we develop a model called WheezeD which, to the best of our knowledge, represents the first step towards developing a computational model for respiration phased based wheeze detection. WheezeD has two components, first, we develop an algorithm to detect respiration phase from audio data. We, then transform the audio into 2-D spectro-temporal image and develop a convolutional neural network (CNN) based wheeze detection model. We evaluate the model performance and compare them to conventional approaches. Experiments on a public dataset show that our model can identify wheezing event with an accuracy of 96.99%, specificity of 97.96%, and sensitivity of 96.08%, which is over 10% improvement in performance compared to the best accuracy reported in the literature by using traditional machine learning models on the same dataset. Moreover, we discuss how WheezeD may be used towards assessment and computation of patient severity.

KEYWORDS

Wheeze, Asthma, Pulmonary Condition, Wheezing Severity, mHealth

ACM Reference Format:
Soujanya Chatterjee, Md. Mahbubur Rahman, Ebrahim Nemati, and Jilong Kuang. 2019. WheezeD: Respiration Phase Based Wheeze Detection Using Acoustic Data From Pulmonary Patients Under Attack. In *Proceedings of EAI PervasiveHealth conference (PervasiveHealth).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Chronic respiratory diseases are described as the chronic diseases of the airways and other structures of the lungs according to the Centers for Disease Control and Prevention (CDC). Some of the common are asthma, chronic obstructive pulmonary disease (COPD), occupational lung disease, chronic bronchitis. An estimated 40 million people in USA have one of the chronic respiratory diseases, more than 50% of them suffer from exacerbations or attacks due to these diseases every year [1, 2]. Respiratory airway obstruction causes change in velocity of air flow during breathing from normal airway into narrowing airway, abnormal breathing sounds are produced, which include wheeze, rhonchus, crackles. Moreover, during an asthma attack or COPD exacerbation, the adventitious sounds may be accompanied by coughing, throat clearing, rapid breathing.

Wheezing is defined as an adventitious, continuous sound which has musical characteristics. It is generated due to rapid breathing through constricted or narrow respiratory airway during COPD or asthma exacerbation. An important aspect of wheezing, as these research works depict, is its impact on patient severity [10, 18, 19]. Severity is the extent of exacerbation of the patient condition under attack which may inform the subsequent intensity of treatment required [21]. Detecting severity and its variation may not only be useful in determining the subsequent level of treatment, but also in identifying the potential causes of variations [7]. Wheeze mostly occurs during an expiratory phase of a respiratory cycle. However, it may also occur during inspiratory phase, or in both of the phases. Severe obstruction of the intra-thoracic airway can be associated with inspiratory wheezes, which are shown to be more severe.

Several works have been done to detect wheeze [15]. However, to the best of our knowledge, this work is the first to present a reliable model for respiratory phase based wheezing detection and its application towards assessment of severity of the pulmonary condition. In this work, we show that it is feasible to detect respiratory phase from acoustic data and subsequently use the phase duration to reliably detect wheezing using a CNN based deep-learning model. By experimenting on a public dataset, named R.A.L.E. [9], we show that the detection model performs better than previous works [12, 17, 20] which use hand-crafted acoustic feature based machine learning model, applied on the same dataset (R.A.L.E. [9].

Related Works: The state-of-the-art method for detecting abnormal or adventitious lung sounds such as wheezing is based on auscultation approach by experienced physicians, using an instrument called stethoscope. This method is non-invasive and simple, however it is heavily dependent on the physical presence of a physician, the experience of the physician, sensitivity and variability of human auditory system [14], presence of noise in the internal or external environment, technical specifications and response efficiency of stethoscopes [4]. The following survey paper summarizes some interesting previous works done on wheezing detection[15]. Few of

*Pursued this work while interning with Samsung Research America Summer 2018

PervasiveHealthEAI 2019, May 20-21
Copyright © 2019 EAI
DOI 10.4108/eai.20-5-2019.2283516

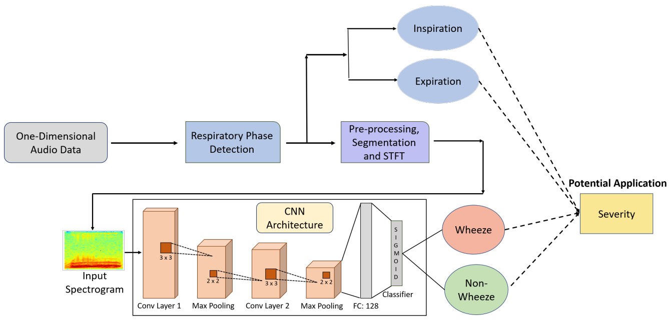

Figure 1: Wheeze detection model architecture

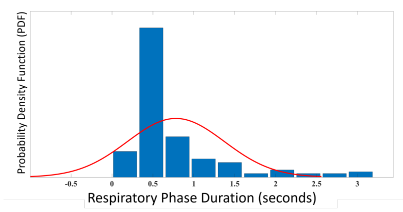

Figure 2: Distribution of respiratory phase duration.

them attempted at wheezing detection against other adventitious sounds. However, we attempt at distinguishing wheezing sounds against non-wheezing sounds which include cough, throat-clearing, regular breathing sounds. Prior works show sufficient evidence about how irregular respiratory pattern in association with the presence of wheezing sound may contribute to the severity. This work shows that high intensity wheezing during both inspiratory and expiratory phase, the duration of wheezing during each phase lead to lower peak expiratory flow rate [19]. High severity may be caused by severe obstruction of the intrathoracic lower airway or upper airways obstruction, which can be manifested by wheezing during the inspiratory phase [18].

2 DATASET DESCRIPTION

We obtain the *wheezing* sound data from R.A.L.E repository [9]. The data set consist of 26 lung sound recordings containing *wheezing* sound from participants with age ranging from newborn to 76 years old. All of these sounds were recorded at $10.240KHz$. R.A.L.E dataset is used as positive class examples for wheezing model. For the negative class examples (cough, throat clearing), we use carefully labeled audio data collected from 21 subjects using Samsung Gear S3 smartwatch ($16KHz$) described in [13, 16]. Moreover, we use an internet repository [3] to obtain throat clearing and additional cough sounds ($44.1KHz$). We use those raw audio samples as input to the model architecture mentioned in Figure 1. It is important to note that we handle the audio samples captured from several sources with varying sampling rate.

3 RESPIRATION PHASE DETECTION

Respiratory phase contains valuable information that can potentially improve the detection accuracy of wheeze events and prediction of the severity. For example, *Wheezing* during inspiratory phase indicates higher level of severity compared to when it happens in expiratory phase [18]. Moreover, severity further worsens when it happens in both phases which is known as biphasic wheeze. In the R.A.L.E dataset [9], we observe that 40% of the patients were having wheeze *during exhalation*, 14% were having wheeze *during inhalation*, and 46% of them were having wheeze in *both phases*. It indicates that dataset used in this modeling covers wheezes in all spectrum of respiratory phases. Moreover, most of the patients were *under pulmonary attack*, which makes the dataset *more valuable and useful for deeper analysis*, such as, the analysis of the respiratory phases in which wheezes are occurring. We utilize this opportunity for the first time to analyze and incorporate the respiratory phase information in modeling *wheeze* events.

We identify respiratory phases by applying appropriate signal processing steps on acoustic data. We observe that respiratory phases (i.e., inhalation and exhalation) can be identified by the envelope of the audio signal in time-domain. Therefore, we apply an envelope detection algorithm to automatically detect the start and end of each respiratory phase. The algorithm, first, detects the peaks and applies cubic spline curve fitting model on the peaks. Since the audio sampling frequency and the variation of the audio signal in time domain is usually very high, we again apply the same envelop detection algorithm on the peaks detected in the first step. We, then, apply third order low-pass Butterworth filter with cut-off frequency of $1.66Hz$ to further smooth the envelop signal. Finally, we apply respiratory phase detection algorithm described in [16] to identify each inhalation and exhalation phase of a respiratory cycle. Several of the identified respiratory phases are verified by listening the audio to ensure that the phases are correctly detected. Figure 2 shows the distribution of the respiratory phase durations. We observe that the most dominant phase duration is $0.5s$ which informs the window size of the deep-learning model described later in this paper. This is how our approach incorporates domain information of a pulmonary attack into the wheeze detection model.

3.1 Processing Acoustic Data

The primary inputs to the model are the images of the sound data in the form of spectrogram which graphically express the sound frequency components and the corresponding time location of their occurrence. Spectrograms from different sources of sound recording may have completely different characteristics, for example, sampling rate may differ across sound sources, producing spectrograms of different spectral resolution. A major challenge prior developing a model is to ensure that the input spectrograms comprise of the same characteristics in the spectro-temporal domain. In order to address this, we adopt the sample rate normalization algorithm from this work [17]. If the sampling rate of the input signal is more than $9KHz$, then the input signal is down-sampled to $9KHz$, in this way we ensure the uniformity of the input signal characteristics.

3.2 Acoustic Data Segmentation

We segment the input spectrogram images (computed using Short-Time Fourier Transformation (STFT) [5]) based on the respiration phase distribution that we obtain from the wheeze dataset (section 3.1). We consider the following candidate window sizes, minimum ($0.25s$), median ($0.5s$), 75th percentile ($0.85s$) and 80th percentile

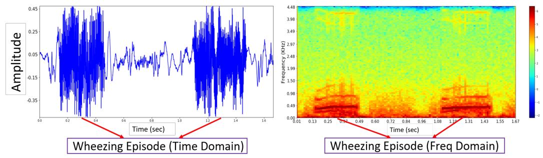

Figure 3: On the left, Time domain 1-D wheezing audio signal. On the right, Spectrogram (2-D image) of the wheezing audio data.

(1s) of the window size distribution. We do not consider window sizes of more than 1s in order to avoid overfitting. We describe the number of sample sizes for each segmentation window later in section 5.2 and table 1. We classify each segment as *"wheeze"* and *"non-wheeze"* events.

4 UNDERSTANDING SPECTROGRAM SIGNATURES OF PULMONARY SOUNDS

In this section, we visualize some of the salient features of the pulmonary sounds from their corresponding spectrograms. Wheezing sound, are of musical and continuous nature, their presence demonstrate a typical signature in the spectrogram, with a continuous near-horizontal lines representing the time interval of the main frequency and the presence of other lines above represents the frequency spectra that compose the harmonic frequency of the main frequency. This unique signature can be observed in figure 3b, where wheeze event occur during inhalation phase. Cough and throat-clearing being a discontinuous sound, characterized by short explosive bursts, having relatively higher amplitude, when compressed in a short-term interval usually generates vertical lines in a spectrogram. The unique spectral characteristics of different pulmonary sounds are captured in their corresponding spectrograms.

5 EXPERIMENTAL METHODS

5.1 Network Architecture

We use Convolutional Neural Network to classify audio data. Our architecture refers to the design of well known network designs like VGG16 [8], LeNet-5 [11]. After several experimentation (refer to section 5.2), our final network consists of 2 convolutional layers, 1 fully connected layer and a sigmoid classification layer. First layer has 16 rectified linear units (ReLU) with filter size of 3 x 3, 2 x 2 max-pooling layer, with stride 1. Second layer has 16 rectified linear units with filter size 5 x 5, 2 x 2 max-pooling layer, with stride 1. The convolutional layers are followed by 1 fully connected layer with 128 neurons. This layer employs dropout regularization of 0.25 in order to reduce overfitting. The sigmoid classification layer which classifies as either a *"wheeze"* event or *"non-wheeze"* event.

We develop the model using the Python programming language, Keras 2.0 deep learning library. The training and testing experiments are performed on 8 Tesla $M40$ $24GB$ GPUs with CUDA toolkit 9.1.

5.2 Experimental Results

We conduct several experiments with the aim of finding the following important hyper-parameters of the model, the optimal window size for classification, number of convolutional layers, filter size. We evaluate the performance on the validation set, in order to demonstrate the effectiveness.

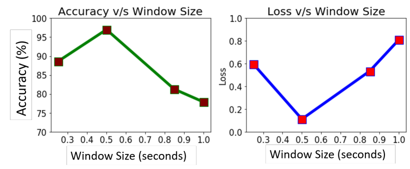

Figure 4: On the left, interestingly, $0.5s$ classification window size produces the highest model accuracy. One the right, we see that the model loss is the least with $0.5s$ classification window size

width=0.5center

Window Size	Sample Size			
	Training	Validation	Test	Total
0.25s	655	280	233	1168
0.5s	504	216	180	900
0.85s	305	132	109	546
1s	275	119	98	492

Table 1

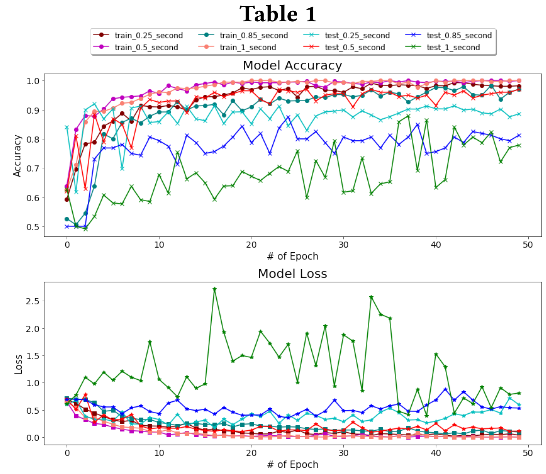

Figure 5: Shows the model accuracy and loss plots for different window of classification. Segmentation window w_i = $0.5s$ performs the best. The loss plot shows that the model experiences significant overfitting when $w_i = 1s$

Obtaining the Optimal window: In order to find the optimal window W, we experiment with a few candidate windows w_i = [$0.25s$, $0.5s$, $0.85s$, $1s$] defined by the inhalation and exhalation duration in section 3.4. In each case, the input data is segmented with the given window size w_i and then the network is trained and validated. We note that the number of training, validation and test samples obtained from the data, depends upon the window size used for segmentation (refer Table 1). We also note that the number of samples are good enough to train a relatively smaller deep-learning network (as in [6]).

$W = 0.5s$ attains an accuracy of 96.999% (figure 4). This can be explained by the fact that as we observe from the dataset, most of the wheezing occur during inhalation or exhalation with a duration of around $0.5s$. If we reduce the window of classification to $0.25s$,

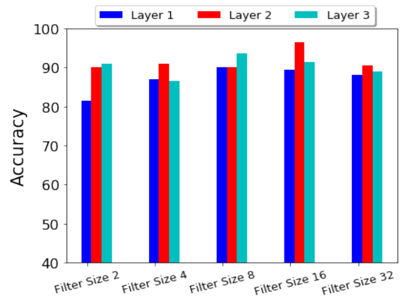

Figure 6: A CNN of 2 layers, 16 filters each performs the best in terms of accuracy

accuracy drops (to 88.588%), we assume that if the window length is too short, then input will not have the discriminative information needed for classification. With further increase in the window length to 0.85*s* (81.249%) and 1*s* (77.885%), which suggest that the model is prone to the risk of heavy overfitting figure shown in figure 5. According to these results, we determine the value of W to be 0.5s.

Number of Convolution layers and filters: We consider 1, 2, and 3 as candidates for the number of layers and 2, 4, 8, 16, 32 for the number of filters. We note from figure 6, that a CNN of 2 layers, each having 16 filters performs the best with accuracy 96.999%. The dropout regularization works well to reduce model overfitting when the model has 2 layers, however introduction of a third layer leads to significant overfitting into the model. From these experimental results, we make an informed decision to limit our model to two layers.

5.3 Model Performance

The performance of the model is a significant improvement over previous detection algorithms developed using the same dataset, which we use as basis of comparison. When evaluated on the test data (completely unseen during model training and validation), we report an overall model accuracy of 96.99%, which is better than that reported in [20] (82.1%), [17] (84.82%).The model obtains a specificity or true negative rate of 97.96%, and sensitivity or true positive rate of 96.08%, better than that reported in [20] (81.5 ± 10% sensitivity and 82.6 ± 7% specificity), [17] (86.1% sensitivity and an 82.5% specificity).

6 LIMITATION, FUTURE WORK AND CONCLUSION

Following are the few limitations which may lead to interesting research opportunities. First, we use limited dataset of wheezing sounds in this work, which in turn restricts the model to two convolutional layers. In future, large amount of data can be collected in order to build deeper and more robust networks. Moreover, we did not explore LSTM based model, which we believe may improve the detection taking into consideration the temporal signature of wheeze in the acoustic data. Second, the detection method will help us collect long term longitudinal wheeze data from patients, which may be useful in assessing triggers of asthma or COPD exacerbation in the wild. Our hope is that, respiratory phase based wheezing detection, like WheezeD can be utilized towards assessment of severity. For instance, severity may be associated with the wheezing duration, rate and diurnal pattern (for e.g., wheezing at night time) [22]. Moreover, severity can be determined by the greater level of obstruction of upper airways, and wheezing during inspiration is a defined surrogate for that [18]. Hence wheezing during inspiration phase may be assessed to be more severe than during expiration. Interestingly, we note that WheezeD presents the fundamental attributes required (respiration phase detection and wheezing detection with much better performance than existing works) to assess wheezing severity.

REFERENCES

[1] Accessed Aug, 2018. *CDC Asthma.* https://www.cdc.gov/nchs/fastats/asthma.htm
[2] Accessed Aug, 2018. *CDC COPD.* https://www.cdc.gov/nchs/fastats/copd.htm
[3] Accessed Aug, 2018. *SoundSnap.* https://www.soundsnap.com/
[4] Manuel Abella, John Formolo, and David G Penney. 1992. Comparison of the acoustic properties of six popular stethoscopes. *The Journal of the Acoustical Society of America* 91, 4 (1992), 2224–2228.
[5] Jont B Allen and Lawrence R Rabiner. 1977. A unified approach to short-time Fourier analysis and synthesis. *Proc. IEEE* 65, 11 (1977), 1558–1564.
[6] Justice Amoh and Kofi Odame. 2015. DeepCough: A deep convolutional neural network in a wearable cough detection system. In *Biomedical Circuits and Systems Conference (BioCAS), 2015 IEEE.* IEEE, 1–4.
[7] Henry Gong et al. 1990. Wheezing and asthma. (1990).
[8] Song Han, Huizi Mao, and William J Dally. 2015. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. *arXiv preprint arXiv:1510.00149* (2015).
[9] PixSoft Inc. Accessed Aug, 2018. *R.A.L.E Lung Sound.* http://www.rale.ca/
[10] Richard S Irwin, Peter J Barnes, FRS Deputy, and Helen Hollingsworth. 2013. Evaluation of wheezing illnesses other than asthma in adults. *UpToDate. Waltham: UpToDate* (2013).
[11] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (1998), 2278–2324.
[12] P Mayorga, C Druzgalski, RL Morelos, OH Gonzalez, and J Vidales. 2010. Acoustics based assessment of respiratory diseases using GMM classification. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE.* IEEE, 6312–6316.
[13] Ebrahim Nemati, Md Mahbubur Rahman, Viswam Nathan, and Jilong Kuang. 2018. Private Audio-Based Cough Sensing for In-Home Pulmonary Assessment using Mobile Devices. In *EAI International Conference on Body Area Networks.*
[14] Hans Pasterkamp, Steve S Kraman, and George R Wodicka. 1997. Respiratory sounds: advances beyond the stethoscope. *American journal of respiratory and critical care medicine* 156, 3 (1997), 974–987.
[15] Renard Xaviero Adhi Pramono, Stuart Bowyer, and Esther Rodriguez-Villegas. 2017. Automatic adventitious respiratory sound analysis: A systematic review. *PloS one* 12, 5 (2017), e0177926.
[16] Md Mahbubur Rahman, Ebrahim Nemati, Viswam Nathan, and Jilong Kuang. 2018. InstantRR: Instantaneous Respiratory Rate Estimation on Context-aware Mobile Devices. In *EAI International Conference on Body Area Networks.*
[17] RJ Riella, P Nohama, and JM Maia. 2009. Method for automatic detection of wheezing in lung sounds. *Brazilian Journal of Medical and Biological Research* 42, 7 (2009), 674–684.
[18] Malay Sarkar, Irappa Madabhavi, Narasimhalu Niranjan, and Megha Dogra. 2015. Auscultation of the respiratory system. *Annals of thoracic medicine* 10, 3 (2015), 158.
[19] Chang S Shim and M Henry Williams. 1983. Relationship of wheezing to the severity of obstruction in asthma. *Archives of internal medicine* 143, 5 (1983), 890–892.
[20] Germán D Sosa, Angel Cruz-Roa, and Fabio A González. 2015. Automatic detection of wheezes by evaluation of multiple acoustic feature extraction methods and C-weighted SVM. In *10th International Symposium on Medical Information Processing and Analysis*, Vol. 9287. International Society for Optics and Photonics, 928709.
[21] DR Taylor, ED Bateman, LP Boulet, HA Boushey, WW Busse, TB Casale, P Chanez, PL Enright, PG Gibson, JC De Jongste, et al. 2008. A new perspective on concepts of asthma severity and control. *European Respiratory Journal* 32, 3 (2008), 545–554.
[22] M Yumiko, S Kazuko, and N Asako. 2015. Pediatric respiratory severity score (PRESS) for respiratory tract infections in children. *Austin Virol and Retrovirology* 2, 1 (2015), 1009.