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 Nemanti
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

1 INTRODUCTION
Chronic respiratory diseases are described as the chronic diseases of the airways and other structures of the lungs according to the

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

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

PervasiveHealth, 2019, Italy
© 2019 Copyright held by the owner/author(s).
ACM ISBN 978-x-yyyy-xxxx-xx/TY/MM.
https://doi.org/10.1145/nmnnnn.nmnnnn

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
We obtain the wheezing sound data from R.A.L.E repository [9]. The dataset consists 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 10.240Hz. 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, **Wheeze** 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 inhalation, 14% were having wheeze during exhalation, 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. 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 wheezing dataset (section 3.1). We consider the following candidate window sizes, minimum (0.25s), median (0.5s), 75th percentile (0.85s) and 80th percentile.
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.

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,
wheeze in the acoustic data. Second, the detection method will which we use as basis of comparison. When evaluated on the test convolutional layers. In future, large amount of data can be collected acerbation in the wild. Our hope is that, respiratory phase based help us collect long term longitudinal wheeze data from patients, did not explore LSTM based model, which we believe may improve research opportunities. First, we use limited dataset of wheezing limitations which may lead to interesting following are the few limitations which may lead to interesting 6 LIMITATION, FUTURE WORK AND CONCLUSION Following are the few limitations which may lead to interesting wheezing detection, like WheezED can be utilized to 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