Diagnosis of Pneumonia From Sounds Collected Using Low Cost Cell Phones

Insu Song
School of Business/IT
James Cook University Australia
Singapore Campus, Singapore
insu.song@jcu.edu.au

***Abstract*—Respiratory diseases, such as pneumonia, cold, flu, and bronchitis, are still the leading causes of child mortality in the world. One solution for alleviating this problem is developing affordable respiratory-health assessment methods using computerized respiratory-sound analysis. This approach has become an active research area due to the recent developments of electronic recording devices, such as electronic stethoscopes. However, all existing methods require specialized equipment, which can be operated only by trained medical personals. We develop a low-cost cell phone-based rapid diagnosis method for respiratory health problems. A total of 367 breath sounds are collected from children's hospitals in order to develop accurate diagnosis models and evaluation. An extensive analysis is performed on the breath sounds. Statistically significance features are selected for each age group using ANOVA from 1197 acoustic features. The model is evaluated on a binary classification task: pneumonia vs. non-pneumonia. The results showed that the proposed method was able to effectively classify pneumonia even in the presence of environmental noises. The method achieved 91.98% accuracy with 92.06% sensitivity and 90.68% specificity. The results indicate that breath sounds recorded using low-cost mobile devices can be used to detect pneumonia effectively.**

Keywords—Mobile diagnosis, Rapid-diagnosis, Respiratory Problem, Lung Sound, Pneumonia, KNN, SVM

I. Introduction

Acute respiratory infections are still one of the leading causes of child mortality in the world. Reports state that around 28-34% of all child mortality under the age of 5 was due to pneumonia and 95% of the total mortality, was from developing countries [1-3]. It is also the major cause of death in low income countries [1]. In this paper our aim is to develop a diagnostic system for pneumonia which can be used by parents of children, in particular by the underserved population. Due to rapid penetration of mobile phones, this is now possible since mobile phones can be used to record high quality breath sounds, which can be either processed on the mobile phone or sent to a remote server for a further analysis [3-6]. However, almost all computerized respiratory sound analysis methods (e.g., [7-10]) were based on the data that were collected by using special sensors attached to the body. This is not only invasive , but also requires a trained personal to properly acquire the data. Currently, no known computerized diagnosis methods available for pneumonia using breathing sounds collected by hand-held microphones (i.e. non-contact-based methods). Therefore, there is a strong need for such a method: whether it would be possible to provide early diagnosis of pneumonia by using breath sounds collected by using non-contact-based methods. Existing literature strongly support that this is possible. For instance, Gupta et al. [11] found that fast breathing was the most useful sign for pneumonia in all age groups in a study of 222 children with pneumonia. World Health Organization (WHO) has also defined pneumonia solely based on the visual inspection and timing of the respiratory rate [12].

The main contributions of this paper are an analysis of breath sounds collected at three children's hospitals in Bangladesh using 18 acoustic feature extraction methods for the diagnosis of pneumonia. We identify features that have high diagnostic power for pneumonia using statistical analysis. Two predictive models are developed using machine learning approaches and evaluated using real data. The method achieved 91.98% accuracy with 92.06% sensitivity and 90.68% specificity. This is more reliable than medical doctors. The results indicate that breath sounds recorded using low-cost cell phones can be used to detect pneumonia effectively. In our best knowledge, this is the first report on the analysis of non-contact-based breathing sounds for the diagnosis of pneumonia.

The paper is organized as follows. Section II discusses related works and select relevant acoustic features for the analysis of breath sounds. Section III illustrates the analysis of breath sounds using the selected acoustic features. Section IV describes the data sets that are used to test diagnostic powers and prediction models for pneumonia. Section V described our methodology of evaluating the diagnostic powers of 1197 acoustic features and the development of two predictive models. Section VI presents the performance results of predictive models. Section VII discuss the analysis results and conclude the paper in Section VIII.

II. Computerized Respiratory Sound Analysis

This work is based on computerized respiratory sound analysis and diagnostic feature extraction. In this section, we overview the core techniques used on respiratory sound analysis.

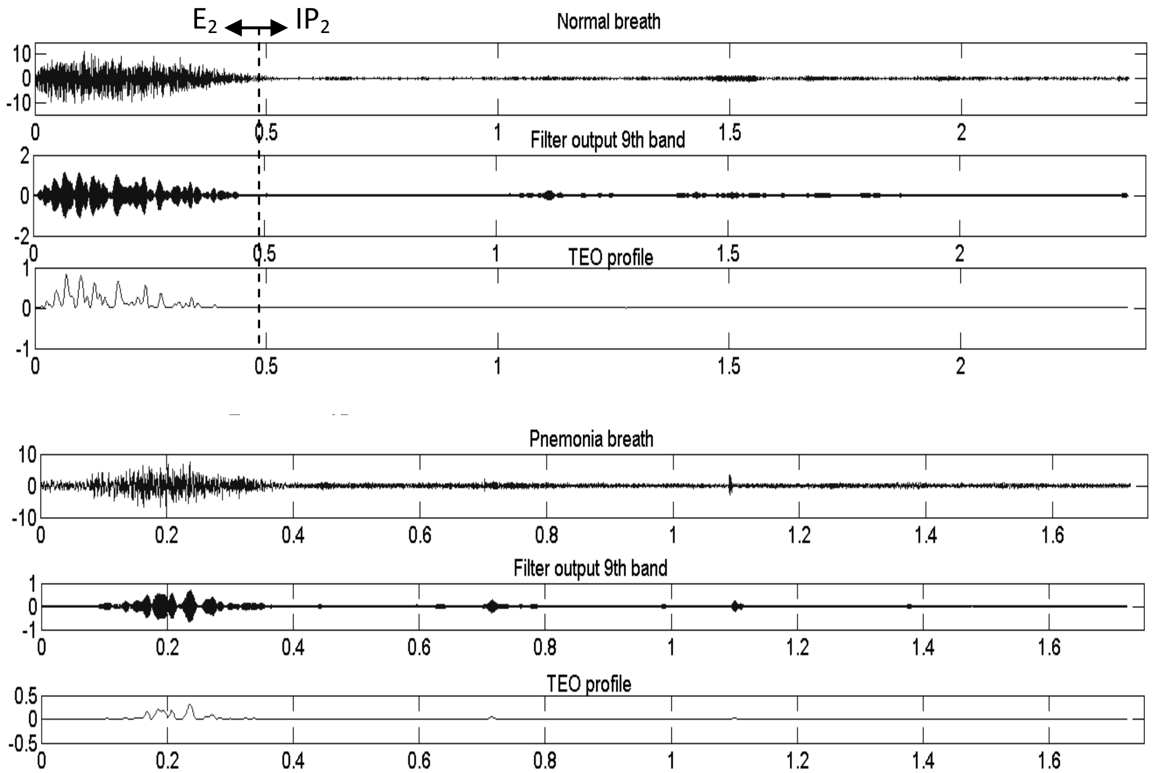


Fig.1. Illustration of two breath sounds and the corresponding TEO-CB-Auto-Env features for 9th critical band for normal breath and pneumonia breath. E denotes the exhalation segment and IP denotes the inhalation segment of breath sound.

Recently many computerized respiratory sound analysis methods have been proposed. Most frequently used sounds are tracheal sounds. They have been frequently used in diagnosing apnea. Kulkas et al. [13] proposed three methods for the compressed tracheal breath sound analysis. The three methods were based on absolute local maximum sound amplitude, local range of maximum sound amplitude and relative range of maximum sound amplitude and achieved significant accuracy in method 2 and 3. Montazeri et al. [9] employed power spectral density, kurtosis and Kartz fractal dimensions to detect Obstructive Sleep Apnea (OSA) on a dataset of 52 participants where 35 participants were pathological and 17 were healthy.

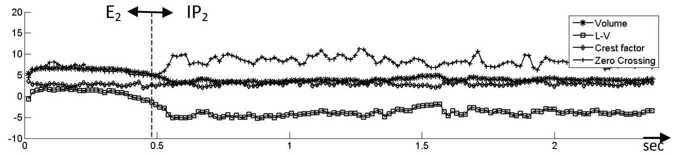
Besides diagnosing apnea some researcher tried to detect respiratory phases from the breath sound. Jin et al. [14] proposed a respiratory phase segmentation method based on multi-population genetic algorithm by introducing an evaluation function based on sample entropy (SampEn) and a heterogeneity measure. Later they proposed another method [10] for signal identification and extraction based on instantaneous frequency (IF) analysis for various adventitious sounds (AS) detection. Huq et al. [15] used log-variance to detect the two phases (inspiratory and expiratory phases) of breath from tracheal sound.

Some researchers used respiratory sounds to diagnose chronic obstructive pulmonary disease (COPD) [16, 17]. Hashemi et al. [16] extracted prosodic features (e.g., volume, skewness, and Kurtosis) from lung sounds to classify into polyphonic and monophonic wheeze sound types: asthma vs. COPD. Tapliduo et al. [17] employed the combination of continuous wavelet transform (CWT) with third-order spectra for the analysis of wheeze sounds to distinguish between asthma and COPD.

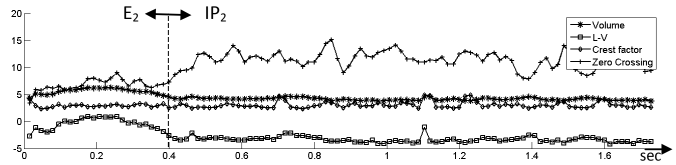
Pneumonia was also diagnosed by using lung sounds [7, 18]. Ono et al. [18] collected lung sounds using an electric condenser microphone attached to back of patients, performed spectral analysis on the sound and found that median and third quartile frequencies were significantly higher in interstitial pneumonia.

However, all of these methods collected breathing sounds by using contact-based method i.e. the equipment used to collect the sound is attached to the body. Most contact-based respiratory sound collection methods require sophisticated apparatuses and special training for use, and therefore, could only be used by trained medical practitioners. Moreover, this type of data collection is uncomfortable for the patient. Furthermore, breath sound collection from young children by contact-based method is difficult due to the restiveness of the children.

On the other hand non-contact-based methods do not introduce discomforts to the patients. Equipment used in non-contact-based methods (e.g. digital voice recorder, mobile phones) are less expensive and requires less supervision during capturing input to the system, therefore this methods are more applicable for untrained users/patients. Moreover, abnormal breathing rate, which is a useful sign in pneumonia, could be easily measured from breath sound collected by non-contact based method using an ordinary digital voice recorder. Hence,



(a)



(b)

Fig.2. Volume, ZCR, CF and LV feature values for (a) Normal Breath and (b) Pneumonia Breath.

the analysis of breath sound collected by non-contact based method could be an interesting and fruitful research area. However, very few studies have been done on the analysis of non-contact-based breath sound (NCBBS). A clinical application on NCBBS was reported by Alasher et al. [19] which used NCBBS mainly for monitoring absence of breath for patients affected in case of anesthesia and sedatives in controlled environment.

III. Acoustic Feature Extraction Methods

In this section, we illustrate 18 types of acoustic features that are analysed for their diagnostic capabilities of pneumonia. The 18 types of acoustic features can be grouped into Prosodic, Spectral, Cepstral, Teager energy operator (TEO), and temporal information features. These features are selected as they are used in the previous studies in the analysis of respiratory sounds as discussed in the previous section.

We use two sample breath sounds collected from children to illustrate the features: one from a child with pneumonia and another from a child without pneumonia.

A. *TEO based features*

Teager devised an algorithm called Teager energy operator (TEO) which used a non-linear energy tracking operator for signal analysis. Teager proposed that the energy of a system is proportional to the amplitude and frequency of the signal produced by the system [20]. TEO for a signal x in discrete time domain n is defined as

$$\psi[x(n)] = x^2(n) - x(n+1)x(n-1)$$

It gives good estimation when the input signal contains sharp changes either in frequency or amplitude.

Several variations of the TEO features are proposed in the literature such as TEO critical-band autocorrelation envelop (TEO-CB-Auto-Env) feature [21]. This feature is computed by first generating 15 Gaussian CBs [21] to generate 15 TEO feature values and then generate autocorrelation envelop for each CB. At last the areas under each envelop are used as the feature values of the frame. Fig. 1 shows the 9th critical band output along with the TEO profile for normal and pneumonia breath.

B. *Prosodic features*

Prosodic features are commonly used features for the analysis of respiratory sounds. These include Fundamental frequency, Formants, Zero Crossing Rate, Volume, Crest Factor, Log-Variance, Skewness, and Kurtosis. Fig. 2 shows volume, ZCR, CF and LV values for a normal breathing sound and pneumonia breathing sound.

Fundamental frequency is the lowest frequency component of a signal. We extracted fundamental frequency by using Roger Jang's toolbox [22].

Formants are the spectral peaks of a sound spectrum. The first three formants are extracted from each segments of a breath cycle by using 13th order LP filter. Reason for using only first 3 formants is that the lower formants can model spoken content [23].

Zero crossing rate (ZCR) is the rate of change of sign of the signal points. There are two approaches for calculating ZCR: (1) does not count "zero positioning" (i.e. signal remains at zero line) as a zero crossing; (2) does count "zero positioning" as a zero crossing. In this paper we extracted ZCR by using the second method.

Volume is the loudness of a sound, sometimes called as energy of a sound, is used to describe the ear's perception of the signal. Volume can be defined in two methods: absolute sum of samples within each frame and log-energy computation. The former is shown below:

$$Volume = \sum_{i=1}^{N} |s_i|$$

where s_i is the ith sample within a frame, and N is the frame size.

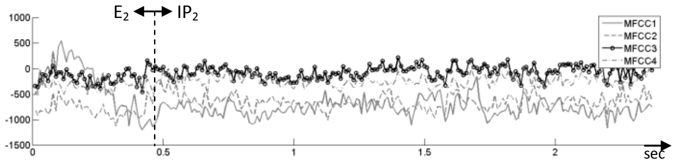
Crest Factor (CF) is calculated as the ratio of peak amplitude (maximum absolute amplitude value in the frame) and RMS (Root-Mean-Square) value (i.e., a square root of the sum of squared signal values) of a signal which is shown below

$$CF = \frac{|s|_{peak}}{s_{RMS}}$$

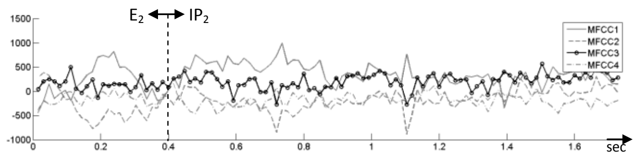
Crest factor gives higher value in case of narrow high peaks and gives lower values for flat signals.

Log-Variance (LV) is computed by taking the logarithm of variance of a sound signal [24] within a frame.

Skewness is a statistical measure of a distribution's degree of deviation from symmetry about the mean. The signal is right skewed if skewness value is positive, if it is negative then the signal is left skewed relative to the mean. Kurtosis is a classical statistical measurement of the "peakedness" of a distribution. A



(a)



(b)

Fig.3. First 4 MFCC feature values for (a) Normal Breath and (b) Pneumonia Breath.

kurtosis of zero is the value for a Gaussian distribution, values less than zero indicate flatter and less peaked distribution, while positive values indicate a narrower and more peaked distribution.

C. Spectral Features

Spectral features are computed by transforming a signal into frequency domain by Fourier transform. We included the following spectral features in our study: Spectral centroid, Spectral flux, spectral entropy, Spectral roll-off, and Power Spectral Density. Spectral centroid, spectral flux, spectral entropy and spectral roll-off are computed by using the toolbox [22]. Power spectral density is computed by using the available functions in MATLAB.

Spectral centroid (SC) determines the frequency area around which most of the signal energy concentrates and is thus closely related to the time-domain Zero Crossing Rate feature. Spectral centroid (SC) is calculated by calculating the mean bin of the power spectrum. The result returned is a number from 0 to 1 that represents at what fraction of the total number of bins this central frequency is.

Spectral flux, which is also called spectral variation, determines changes of spectral energy distribution of two successive frames. It is calculated by first calculating the difference between two consecutive frames' magnitude spectrum bin, then squared the difference, and at last take the sum of the squares.

Spectral entropy is often calculated by normalizing the power spectrum and then transform with the Shannon's information theory function. Spectral entropy permits separation of the contributions from different frequency ranges. For example, using spectral entropy, one can separate the high-frequency contribution (>32 Hz) from the low-frequency contribution (< 32 Hz).

The spectral roll-off point is the fraction of bins in the power spectrum at which 85% of the power is at lower frequencies [25]. This is a measure of the amount of the right-skewness of the power spectrum.

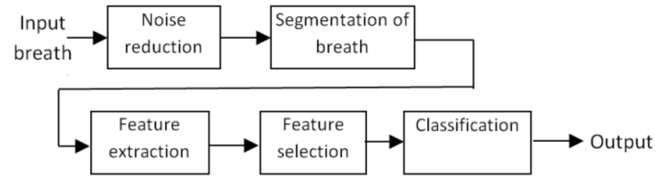


Fig.4. Processing steps of the proposed system.

Power spectral density estimates how the energy of a signal is distributed over different frequencies. We employed Welch method of computing PSD which calculate PSD by firstly partitioning the input signal into segments, compute a modified periodogram of each segment and lastly take the average of the PSD estimates.

D. Cepstral Features

Cepstral features include Cepstral Features, and Mel-scale frequency cepstral coefficients (MFCC). Cepstrum is a Fourier analysis of the logarithmic amplitude spectrum of the signal. It contains information on the rate of change in the different frequency components. This is used for the analysis of the complex dynamics of lung sound and often transformed to MFCC for separation of periodic and harmonic signals.

For speech or speaker recognition, most commonly used acoustic feature is Mel-scale frequency cepstral coefficients (MFCC). MFCC is computed by the transformation of logarithmically compressed filter output. These filters are derived from a triangular filter bank that processes the discrete Fourier transform of a speech signal [26]. In our system we extracted first 13 MFCC features by using 20 triangular filters since most of the respiratory sound analysis systems used these parameters [27-29]. The first four MFCC features are shown in Fig. 3 for normal and pneumonia breath sounds.

E. Delta and Delta-Delta Coefficients

To capture the temporal information among the neighbouring frames we calculate first-order (Delta) and second-order (Delta-Delta) coefficients for some features. The Delta coefficients are computed using linear regression formula:

$$d_t = \sum_{\theta=1}^{\Theta} \theta\left(c_{t+\theta} - c_{t-\theta}\right) \Big/ 2\sum_{\theta=1}^{\Theta} \theta^2$$

where d_t is the delta coefficient at time t, c_t is the feature coefficient, and Θ is the window size which is 10 in our system. The Delta-Delta coefficients are computed using linear regression of Delta features. We computed Delta and Delta-Delta statistics for all prosodic features except breath duration, fundamental frequency and formants, all cepstral features, and TEO features.

IV. BREATH SOUND DATABASE FORMULATION

We collected respiratory sounds from 376 children in three pediatric clinics in Bangladesh using a digital stereo voice recorder (SONY ICD-UX523F) where breathing sound was

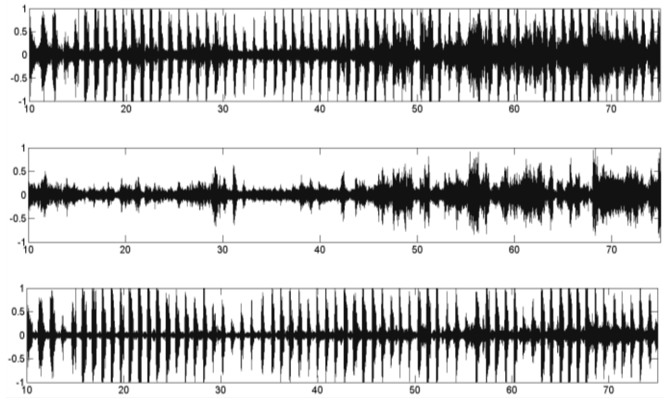


Fig.5. Illustration of the noise reduction process: (a) left-channel input with breath sound, (b) right-channel input, and (c) the breath sound after the noise reduction process.

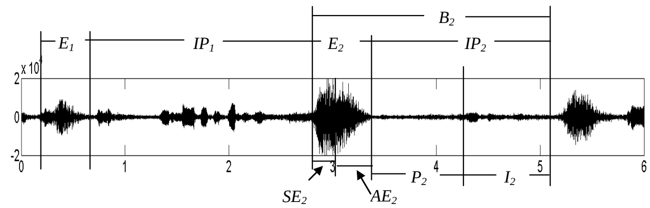


Fig.6. Illustration of the seven segments of breath sound cycles.

recorded using the left-channel. During the collection of the breath sounds, parents of the children were asked about general well-being questions (e.g. energetic and sluggish). Information on the respiratory health related conditions such as having cough, flu, pneumonia, were provided by the medical doctors in the clinic, which were verified by repeated visits of the patients and progress of the symptoms. Written consents were obtained from the parents for research purpose (under the ethics clearance of James Cook University, Australia). Among the 376 children (age range of 0.3 to 5 years, gender distribution of male 212 and female 164), 153 had pneumonia. The sound was recorded by placing the microphone near nose or mouth (if nose was blocked) for at least 1 minute.

The Audio sampling rate was 44.1 KHz with analogue to digital conversion with 16-bit quantization. Recorded sounds were normalized and down sampled to 8 KHz to simulate the telecommunications channels of PSTN (public switched network) and VoIP (voice over Internet protocol).

V. METHODOLOGY

Fig. 4 shows the overall processing steps of our system. First the system takes a breathing sound as the input by using a digital voice recorder. For this purpose we used a stereo digital voice recorder where breathing sound was captured by the left channel and the environmental noise was captured by the right channel. In the noise reduction step we subtracted the environmental noise from the breathing sound to reduce noise. Next we annotated the breathing sound into different segments such as exhalation and inhalation. Different acoustic features were extracted from different segments of the breath in feature extraction step. Next we performed a statistical analysis on the extracted features and selected statistically significant features. Finally the decision was taken by a classifier whether the input breath sound is from a pneumonia patient or not.

A. Noise Reduction

The breath sounds were collected by using dual microphone voice recorders where the breathing sound was recorded by using the left channel. This indicates that the right channel captured mainly the environmental noises. We employed a simple noise reduction technique: subtract the right channel from the left channel to get noise reduced breathing sounds. These types of methods were shown to be suitable for noise reduction [30]. Fig. 5 shows one input signal before and after the noise reduction.

B. Feature Extraction

To extract different features we first annotated a breath cycle into 7 segments as shown in Fig. 6. A breath cycle B_i is first divided into two segments: expiration times E_i (i.e., the time during which the patient was exhaling), and inspiration times IP_i (i.e., the time during which the patient was inhaling). Ei is further divided into spontaneous expirations SE_i and active expirations AE_i. Similarly inspiration times IP_i is further divided into pauses Pi and inspirations I_i. For the time being these annotations are done manually. After obtaining the 7 segments of a breath, different features are extracted from each segment. For computing different features we extracted features from the first three breaths by using 32ms of frame size with 50% overlap and calculated simple statistics like mean and standard deviation to represent the features.

For each segment, we extracted the features describe in Section III. Along with those features we extracted another prosodic feature: breathing duration. We extracted durations for the following five segments: B_i, E_i, IP_i, SE_i and AE_i. Along with these durations we also computed the rate of change of E_i and IP_i which are calculated as $1/(|E_i\text{-}1 - E_i|)$ and $1/(|IP_i\text{-}1 - IP_i|)$ respectively.

C. Statistical Analysis and Feature Selection

A total of 18 types of features were computed with 1197 coefficients which include mean, standard deviation and for some features delta and delta-delta statistics. To reduce the dimensionality of the feature space we performed statistical analysis on the extracted feature values and chose the statistically significant features. For this purpose we first grouped the feature values into age groups: 3-12 months, 13-24 months, 25-36 months, 37-48 months and 49-60 months. For each group we checked the normality of the feature coefficients using Kolmogorov-Smirnov test ($p>0.05$) where we found all the feature categories were normally distributed. Then we performed one-way analysis of variance (ANOVA) within each age group for each feature coefficients values by using MATLAB. Features that were significant at $p<0.05$ in ANOVA test were then selected for classification, and other features were discarded. By the ANOVA test 302, 70, 210, 131 and 152

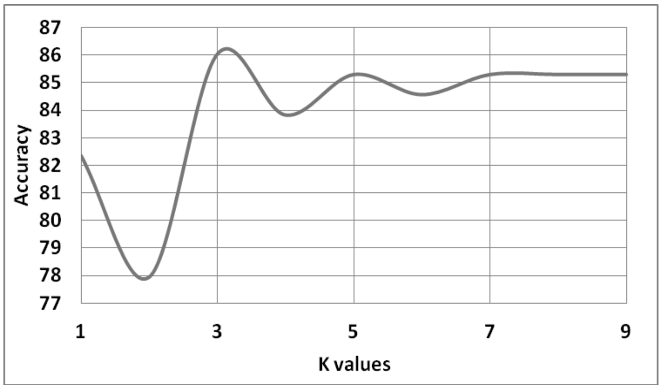


Fig.7. Plot of the accuracy of age group 3-12 months for different values of 'K' of K-NN classifier.

features were selected for 3-12 months, 13-24 months, 25-36 months, 37-48 months and 49-60 months, respectively.

VI. Experimental Results

For classification we employed two classifiers: K-Nearest Neighbour (K-NN) and Support vector machine (SVM) by using WEKA [31]. Each classifier was applied on each age group separately. To evaluate the classification performance, we used the Leave-one-out cross validation (LOOCV) method since in each age group number of data is not big (e.g. in age group 25-36 months contains 41 data). In case of LOOCV we take one instances out from the dataset of one age group as test data and the rest of the dataset is used for training. This is repeated for all the instances in the dataset and the final result is taken by averaging all the result. For the training purpose of SVM, sequential minimal optimization (SMO) algorithm was utilized [32]. For K-NN we chose an optimum value for K experimentally for one age group and applied to all the age groups. Fig. 7 shows the accuracy values of the K-NN classifier for the age group (3-12 months) over different K values. From this experiment we chose the best K value as 3.

Correct classification of pneumonia effected breath and normal breath were measured in terms of sensitivity, specificity and the overall accuracy which are defined as follows.

$$Sensitivity = \frac{TP}{TP + FN} \times 100\%$$

$$Specificity = \frac{TN}{TN + FP} \times 100\%$$

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \times 100\%$$

where True positive (TP) = Number of pneumonia breath is classified as pneumonia breath, False negative (FN) = Number of pneumonia breath classified as normal breath, True negative (TN) = Number of normal breath classified as normal breath, and False positive (FP) = Number of normal breath classified as pneumonia breath.

Classification results of different age groups for the two classifiers are shown in Table I. The average accuracy for K-NN was 90.68% and for SVM was 91.98%. The system achieved specificity of 89.88% and 90.68%, sensitivity of 90.22% and 92.06%, for K-NN and SVM respectively. These results indicate that the breath sounds collected by using non-contact-based method can be used to effectively distinguish between infected and normal child.

VII. Discussion and Analysis of Results

Breath sounds of children of different age contain different properties therefore the feature values of different ages will vary (e.g. breathing rate). Hence instead of analysing all age group in combined, we divided the whole dataset into five age groups and found exciting results in all age groups especially in the age group of 49-60 months. We segmented one breathing sound into 7 segments to allow our system to capture the properties of all the possible cycles of a breath. There could be significant relationships among the different segments of breath cycles and one particular segment could have major contribution in the identification of pneumonia. This analysis is left as a future work. Two classifiers were used to classify the infected breathing sound: a lazy learning classifier (K-NN) and an eager learning classifier (SVM). Both the classifiers gave almost the same results which mean that the selected features by statistical analysis have good descriptive power that do not depend on the efficiency of individual classifier. High sensitivity and specificity with high accuracy indicate that the proposed method of diagnosing pneumonia is very effective on the collected data. These data were collected by persons without any training using ordinary voice recorders and in a real non-controlled environment where lots of ambient noises were present. Hence, the collected data have significant amount of environmental noises which was suppressed in our system by using simple stereo subtraction method, which were found to be very effective. Above discussion proves that the breath sound collected at the mouth of children by ordinary digital voice recorder can be used for diagnosing major respiratory diseases like pneumonia effectively even in the presence of environmental noises.

Table 1. Experimental results over age (months) groups and classifiers.

Age Groups	**K-NN (K=3)**			**SVM**		
	Acc	**Sen**	**Spec**	**Acc**	**Sen**	**Spec**
3to12	86.03%	87%	84.1%	84.56%	87%	79.5%
13to24	78%	85.2%	69.6%	86%	88.9%	82.6%
25to36	92.68%	88.9%	95.7%	92.68	94.4%	91.3%
37to48	96.67%	90%	100%	96.67%	90%	100%
49to60	100%	100%	100%	100%	100%	100%
Avg	**90.68%**	**90.22%**	**89.88%**	**91.98%**	**92.06%**	**90.68%**

[a.] Number of patients in each group: 150 for 3 to 12, 70 for 13 to 24, 41 for 25 to 36, 32 for 37 to 48, and 74 for 49 to 60.

VIII. Conclusions

Existing literature of diagnosing pneumonia works based on the breath sound collected by contact-based method and achieves around 80% accuracy though the signal to noise ratio is high in those data (e.g. [33]). In this paper we built a dataset on pneumonia from paediatric hospitals where breath sounds were collected by non-contact-based method, then extracted different acoustic features from those data and achieved 90.06% accuracy with 92.06% sensitivity and 90.68% specificity. These results indicate that respiratory related major diseases like pneumonia can be diagnosed from the breath sound collected by non-contact-based method. These data can be collected by using any ordinary digital voice recorder such as mobile phones. Penetration of cell phone network and lower cost of mobile sets makes the cell phones available to the people who live in remote places. For those people, mobile phones can be used as a diagnosis tool for different respiratory related diseases since mobile phones can record breathing sounds and can either process it by itself or can send it to the high-computing processors through the cell phone network for programmatically expensive processing. Hence, the proposed system can be used as an early child health intervention for respiratory related disease in the region of limited transportation system [3] and for health social networks [34]. In future, we will develop automatic system for annotating breath sound and test the system for other respiratory related diseases.

ACKNOWLEDGMENT

This project is funded by a grant from the Bill & Melinda Gates Foundation through the Grand Challenges Explorations Initiative (Grant Number: OPP1032125). The authors would like to thank S. A. Rahman for his initial work in data gathering, analysis, formatting, and preparing the paper.

REFERENCES.

[1] O. Ruuskanen, E. Lahti, L. C. Jennings, and D. R. Murdoch, "Viral pneumonia," *The Lancet,* vol. 377, pp. 1264-1275, 2011.

[2] V. Singh and S. Aneja, "Pneumonia – Management in the Developing World," *Paediatric Respiratory Reviews,* vol. 12, pp. 52-59, 2011.

[3] J. Vong and I. Song, *Emerging Technologies for Emerging Markets* vol. 11: Springer, 2015.

[4] I. Song and J. Vong, "Affective Core-Banking Services for Microfinance," in *Computer and Information Science*, ed: Springer, 2013, pp. 91-102.

[5] J. Vong, J. Fang, and I. Song, "Delivering financial services through mobile phone technology: a pilot study on impact of mobile money service on micro–entrepreneurs in rural Cambodia," *International Journal of Information Systems and Change Management,* vol. 6, pp. 177-186, 2012.

[6] I. Song, C. Lui, and J. Vong, "Lowering the interest burden for microfinance," *International Journal of Process Management and Benchmarking,* vol. 4, pp. 213-229, 2014.

[7] L. E. Ellington, R. H. Gilman, J. M. Tielsch, M. Steinhoff, D. Figueroa, S. Rodriguez*, et al.*, "Computerised lung sound analysis to improve the specificity of paediatric pneumonia diagnosis in resource-poor settings: Protocol and methods for an observational study," *BMJ Open*, vol. 2, 2012.

[8] Y. Amimoto, H. Nakano, N. Masumoto, A. Ishimatsu, Y. Arakaki, N. Taba*, et al.*, "Lung sound analysis in a patient with vocal cord dysfunction and bronchial asthma," *Journal of Asthma,* vol. 49, pp. 227-229, 2012.

[9] A. Montazeri and Z. Moussavi, "Obstructive sleep apnea prediction during wakefulness," in *Engineering in Medicine and Biology Society, (EMBC)*, 2011, pp. 773-776.

[10] F. Jin, S. Krishnan, and F. Sattar, "Adventitious sounds identification and extraction using temporal-spectral dominance-based features," *IEEE Transactions on Biomedical Engineering,* vol. 58, pp. 3078-3087, 2011.

[11] D. Gupta, S. Mishra, and P. Chaturvedi, "Fast Breathing in the Diagnosis of Pneumonia—A Reassessment," *Journal of Tropical Pediatrics,* vol. 42, pp. 196-199, August 1, 1996 1996.

[12] T. Puumalainen, B. Quiambao, E. Abucejo-Ladesma, S. Lupisan, T. Heiskanen-Kosma, P. Ruutu*, et al.*, "Clinical case review: A method to improve identification of true clinical and radiographic pneumonia in children meeting the World Health Organization definition for pneumonia," *BMC Infectious Diseases,* vol. 8, p. 95, 2008.

[13] A. Kulkas, E. Rauhala, E. Huupponen, J. Virkkala, M. Tenhunen, A. Saastamoinen*, et al.*, "Detection of compressed tracheal sound patterns with large amplitude variation during sleep," *Medical and Biological Engineering and Computing,* vol. 46, pp. 315-321, 2008.

[14] F. Jin, F. Sattar, and D. Goh, "An acoustical respiratory phase segmentation algorithm using genetic approach," *Medical and Biological Engineering and Computing,* vol. 47, pp. 941-953, 2009.

[15] S. Huq and Z. Moussavi, "Automatic breath phase detection using only tracheal sounds," presented at the International Conference on Engineering in Medicine and Biology Society (EMBC), 2010.

[16] A. Hashemi, H. Arabalibiek, and K. Agin, "Classification of Wheeze Sounds Using Wavelets and Neural Networks," in *International Conference on Biomedical Engineering and Technology*, 2011, pp. 127-131.

[17] S. A. Taplidou and L. J. Hadjileontiadis, "Analysis of wheezes using wavelet higher order spectral features," *IEEE Transactions on Biomedical Engineering,* vol. 57, pp. 1596-1610, 2010.

[18] H. Ono, Y. Taniguchi, K. Shinoda, T. Sakamoto, S. Kudoh, and A. Gemma, "Evaluation of the usefulness of spectral analysis of inspiratory lung sounds recorded with phonopneumography in patients with interstitial pneumonia," *Journal of Nippon Medical School,* vol. 76, pp. 67-75, 2009.

[19] H. Alshaer, G. Fernie, and T. Bradley, "Monitoring of breathing phases using a bioacoustic method in healthy awake subjects," *Journal of Clinical Monitoring and Computing,* pp. 1-10, 2011.

[20] H. M. Teager, "Some observations on oral air-flow during phonation," *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. 28, pp. 599-601, 1980.

[21] G. Zhou, J. H. L. Hansen, and J. F. Kaiser, "Nonlinear feature based classification of speech under stress," *Speech and Audio Processing, IEEE Transactions on,* vol. 9, pp. 201-216, 2001.

[22] R. Jang. (1996). *Audio Processing Toolbox. Available: http://neural.cs.nthu.edu.tw/jang/*.

[23] B. Schuller, A. S. Batliner, D., S. Steidl, T. Vogt, J. Wagner, L. Devillers*, et al.*, "The relevance of feature type for the automatic classification of emotional user states: low level descriptors and functionals," presented at the Interspeech, 2007.

[24] A. Yadollahi and Z. Moussavi, "Formant analysis of breath and snore sounds," presented at the International Conference on EMBC, 2009.

[25] L. S. A. Low, M. C. Maddage, M. Lech, L. B. Sheeber, and N. B. Allen, "Detection of Clinical Depression in Adolescents' Speech During Family Interactions," *IEEE Trans Biomed. Engg.,* vol. 58, pp. 574-586, 2011.

[26] M. Hossan, S. Memon, and M. Gregory, "A novel approach for MFCC feature extraction," presented at the International Conference on Signal Processing and Comm. Sys., 2010.

[27] S. Matos, S. S. Birring, I. D. Pavord, and H. Evans, "Detection of cough signals in continuous audio recordings using hidden Markov models," *IEEE Transactions on Biomedical Engineering,* vol. 53, pp. 1078-1083, 2006.

[28] W. D. Duckitt, S. K. Tuomi, and T. R. Niesler, "Automatic detection, segmentation and assessment of snoring from ambient acoustic data," *Physiological Measurement,* vol. 27, pp. 1047-1057, 2006.

[29] A. S. Karunajeewa, U. R. Abeyratne, and C. Hukins, "Multi-feature snore sound analysis in obstructive sleep apnea–hypopnea syndrome," *Physiological Measurement,* vol. 32, p. 83, 2011.

[30] D. O'Shaughnessy, "Enhancing speech degrated by additive noise or interfering speakers," *Comm. Mag.*, vol. 27, pp. 46-52, 1989.
[31] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *SIGKDD Explor. Newsl.*, vol. 11, pp. 10-18, 2009.
[32] J. C. Platt, "Fast training of support vector machines using sequential minimal optimization," in *Advances in kernel methods*, ed: MIT Press, 1999, pp. 185-208.
[33] A. Gurung, C. G. Scrafford, J. M. Tielsch, O. S. Levine, and W. Checkley, "Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: A systematic review and meta-analysis," *Respiratory Medicine*, vol. 105, pp. 1396-1403, 2011.
[34] I. Song and N. V. Marsh, "Anonymous Indexing of Health Conditions for a Similarity Measure," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 16, pp. 737-744, 2012.