A Computer Assisted Diagnosis System for the Identification/ Auscultation of Pulmonary Pathologies

Macario O. Cordel II* and Joel P. Ilao
Center for Automation Research, De La Salle University, Manila, Philippines

*Coresponding author: macario.cordel@dlsu.edu.ph

ABSTRACT

Statistics show that the primary cause of morbidity and mortality among Filipinos are pulmonary illnesses. These illnesses could have been prevented if detected and treated early. With the physicians' medical knowledge and experience, early detection of possible common pulmonary diseases can be performed using a stethoscope. However, with the current physician-to-population ratio in the country, early detection of respiratory diseases may not be performed on most cases especially in the rural areas, causing even benign cases to lead to mortality. In this paper, we present the development of a system that classifies lung sound for possible pulmonary pathology. Using an electronic stethoscope, lung sounds were collected from healthy individuals and patients with common pulmonary problems for the developed system's training and evaluation. The collected data were pre-processed in order to remove mechanical and other external noises. Using Support Vector Machine (SVM) for modelling and classification, the developed system was able to achieve 100% identification of the normal lung sound from the adventitious lung sound, with an average cross-validation performance of 88%. The developed system, however, has low performance in classifying specific lung sounds, that is, normal vs. crackle vs. wheeze vs. ronchi, with an average accuracy of 61.42% and an average cross-validation performance of 90%.

Keywords: Computer assisted diagnosis, Lung sound enhancement, Lung sound classification, Pattern recognition, Electronic stethoscope, Support Vector Machine (SVM)

BACKGROUND OF THE STUDY AND RELATED WORKS

From 2000 to 2005, the top two leading causes of morbidity among Filipinos, both in rural and urban areas, are classified as respiratory diseases (Department of Health, 2013a) and three (acute lower respiratory tract infection, pneumonia and bronchitis) out of the 10 leading causes of morbidity are on the same classification (Department of Health, 2013b). Moreover, from 2000 to 2006, three out of five main causes of mortality in the Philippines are also respiratory diseases, which are (1) pneumonia, (2) tuberculosis, and (3) chronic lower respiratory tract infection (Department of Health, 2013b). Early detection of these diseases is important for their successful treatment. However, the current doctor-to-patient ratio of 12:10000 (Samaniego, 2011; Duque, 2009) makes early detection of diseases very difficult, if not impossible. Recent technology now allow physiological signals to be conveniently recorded and electronically transmitted, allowing more patients early access to expert advice, and hence, better chances for a successful treatment.

The stethoscope is a basic but very essential equipment used by physicians for diagnosis of cardiovascular and pulmonary diseases. Because special training is required so it can be used to detect abnormal lung sounds, a stethoscope cannot be used in areas where there is no doctor. To make remote diagnosis of lung pathology more convenient for the medical doctor, several studies were made for sound enhancement and automatic identification of pulmonary diseases.

(Varady, 2001; Bai & Lu, 2005; Schmidt, Holst-Hansen, Graff, Toft, & Stuijk, 2007; Yu, Bilberg, & Voss, 2008; Shah & Papadias, 2013; Sakai, Kato, Miyahara, & Kiyasu, 2012; Emmanouilidou, Patil, West, & Elhilali, 2012; Himeshima, Yamashita, Matsunaga, & Miyahara, 2012; Kaya & Elhilali, 2013; Li, Wu, & Du, 2012; Li & Du, 2005).

With the advancement of digital signal processing technology, several signal enhancements and abnormality detection schemes for heart sound signals have been suggested (Varady, 2001; Bai & Lu, 2005; Schmidt et al., 2007), (Yu et al., 2008) that aids in denoising acquired sounds via an electronic stethoscope. However, in chest auscultation for cardiorespiratory sound analysis, collected lung sound signals are accompanied with speech, snore, and other external noises, such as mechanical noise due to friction between the stethoscope and the skin.

Shah and Papadias (2013) were able to separate the heart sound and the lung sound using the Degenerate Unmixing Estimation Technique (DUET). DUET exploits the sparsity of the heart and lung sounds and was first used on speech signals on the assumption that the speech sources in the mixture are disjoint. This blind source scheme for separating the two sound sources, that is heart and lung sounds, was able to recover the cardiorespiratory sound signals with high SNR. However, their work was focused on recovering only the heart sound impaired with the lung sound.

Detection of adventitious lung sounds was described in Sakai et al. (2012), Emmanouilidou et al. (2012) and Himeshima et al. (2012). In Sakai et al. (2012), adventitious lung sounds were detected from low-quality auscultation signals by extracting lung sound components via sparse representation. This scheme works on the idea that noise cannot be represented sparsely by any base. That is, the system extracts the adventitious sounds by identifying the non-zero coefficients of the basis functions that synthesize the low-quality lung sound. This system consistently achieves an average precision of 85% regardless of the noise level. However, the noise in this problem is assumed to be stationary.

Another approach that detects abnormal lung sounds impaired with non-stationary noise is presented by Himeshima et al. (2012).

This work also considers, in addition to spectral content, the duration of the abnormal sounds due to the similarity in spectral content of both the noise and the adventitious sound, especially the wheezes. The work establishes that the noise duration and adventitious sound duration have different normal and gamma distributions. Using gamma distribution, the classification performance reaches up to 90.0%.

Other previous works (Guntupalli, Alapat, Bandi, & Kushnir, 2008; Waitmann, Clarkson, Barwise, & King, 2000; Kahya, Yeginer, & Bilgic, 2006; Lu & Bahoura, 2008; Kandaswamy, Kumar, Ramanathan, Jayaraman, & Malmurugan, 2004; Riella, Nohama, & Maia, 2009) have focused on identifying adventitious sounds, particularly wheezes and crackles. These solutions, similar to Himeshima et al., (2012), capture the spectral and temporal details of the sounds via frequency analysis (Guntupalli et al., 2008; Waitmann et al., 2000), or time-frequency and wavelet analysis (Kahya et al., 2006; Lu & Bahoura, 2008; and Kandaswamy et al., 2004).

Another scheme (Emmanouilidou et al., 2012) was proposed, that mimics the human auditory system in detecting abnormal lung sounds in paediatric auscultation recordings under noisy conditions. This system applies a series of filters to model the cochlear filters, inner hair cell potentials, phase locking, and cortical neurons. Support Vector Machine (SVM) was then used to separate the patterns in cortical responses. Despite not having to subject the signals to a denoising phase, the system was able to achieve sensitivity equal to 89.44% and specificity equal to 80.50%.

A study by Kaya & Elhilali (2013) was performed to determine the classification accuracy of using spectro-temporal features to detect abnormal sounds in noisy bio-signals. The scheme employs recursive tracking of temporal patterns in the lung signal using Kalman filtering. On per segment (inhale-exhale) basis, it achieves 71% normal detection and 92.5% abnormal detection. Considering the events or noise that occur in each segment, 95% normal identification was achieved when the noise contribution is due to the stethoscope movement only. Ambient sound impairment to the signal results in 85.71% normal lung sound identification. The study only considers classification as a normal lung sound or an abnormal lung sound, regardless of whether the abnormal lung sound is a wheeze, crackle, or a rhonchus.

In summary, initial work on lung sound feature extraction and classification can already discriminate a normal lung sound from adventitious or abnormal lung sounds. Technology-assisted medical services, such as diagnosis of lung sounds taken via electronic stethoscope, can help mitigate the lack of qualified medical practitioners in maintaining community health care services especially in far-flung areas, if properly implemented.

Thus, this work aims to develop a software system, with the help of an electronic stethoscope as the acquisition device, which identifies lung sounds, that is, normal lung sound, wheezes, crackles and ronchi, for the identification of possible common pulmonary pathologies. Specifically, this study aims to: (1) Collect chest sounds from healthy people and patients with common respiratory diseases using commercially available electronic stethoscope, (2) Determine a filtering scheme that can separate enough important information of lung sounds from other chest sounds, for example, heart beat sound and external noise, (3) Manually sort and label the lung sounds into normal, wheeze, and crackle sound, with the help of qualified physicians, (4) Utilize both the time and frequency domain-based signal enhancement for feature extraction, (5) determine the machine learning algorithm that can best classify the lung sound using the patterns, and (6) Evaluate the performance of the system by comparing the labels of the expert on the lung sound and the automatic identification/labelling of the developed system.

THE NORMAL AND ADVENTITIOUS LUNG SOUNDS

Breathing consists of two respiratory phases: inspiration and expiration. The entire breathing cycle needs to be observed since minimal changes in breathing rate, depth, and timing could possibly imply a disease. Breath sounds may be divided into three categories: (1) Normal, (2) Abnormal, and (3) Adventitious. A normal vesicular lung sound typically has louder and longer inspiratory sound than expiratory sound. Abnormal lung sounds are often correlated with the absences of breath sound or the presence of breath sound in areas where it is not generally heard. "Adventitious" sounds, which are superimposed on normal lung sounds, in certain circumstances, usually denote a disease. There are three common types of adventitious lung sounds, namely, (1) crackles, (2) wheezes, and (3) rhonchi.

Normal Lung Sound

Generally, lung sound frequency ranges from 50 Hz to 2500 Hz, except for tracheal sounds which sometimes reach up to 4 kHz (Reichert, Gass, Brandt, & Andres, 2008). Lung sounds can be heard along the areas of the trachea, bronchioles, at the back of the chest between the scapula, and mostly throughout the lung field. Lung sounds can be classified according to the region where they are produced/observed. Vesicular breath sound is heard in most parts of the lungs and constitutes the majority of normal lung sounds. It is characterized as soft and low-pitched. Inspiratory sound is louder with respect to expiratory sound without a pause between them.

Crackles

Crackles are discontinuous, explosive, popping sounds that originate within the airways and are most commonly heard during inspiratory phase rather than the expiratory phase. According to Forgac's theory (Piirila & Sovijarvi, 1995), during inspiration, a gas pressure is developed across the airways which collapse during expiration. The crackling sound is produced when a closed airway suddenly opens during inspiration or closes during expiration. Each abrupt opening or closing of an airway is represented by a single crackle.

Table 1. Summary of lung sound characteristics

Lung sounds	Frequency Range	Temporal Features
normal	Low pitched, 50 Hz to 1000 Hz, up to 2500 Hz	Soft, longer and loader inspiration over expiration
crackles	Low (fine) or high pitched (coarse) 100 Hz-2000 Hz	Duration (inspiration + expiration) < 20 ms
wheezes	100 Hz to 1 kHz	80 ms < Duration < 250 ms
ronchi	< 300 Hz	Duration > 100 ms

The crackle mechanism can have two probabilities: (1) An obstructed area called the distal airway suddenly opens and the pressures on either side of the obstruction (bronchiole and alveolus) suddenly equalize resulting in transient, sharp vibrations in the airway wall, and (2) the bubbling of air through secretions in the trachea and bronchi or so-called the larger airways. These two types of mechanisms classify the crackle respectively as either fine or coarse. Fine crackles are calmer, higher-pitched, and have shorter duration. Coarse crackles are relatively louder, lower-pitched, and have longer duration than fine crackles.

Crackles are generally known for their explosive sound, usually denoting a pulmonary disorder (Reichert et al., 2008). They are generated by the opening of an abnormally-closed airway during inspiration or during the closing in the expiration (Reichert et al., 2008; Piirila & Sovijarvi, 1995). Crackles duration are generally lower than 20 ms and the frequency range is between 100-2000 Hz (Reichert et al., 2008; Sahgal, 2011).

Wheezes

Wheezes are continuous musical tones or sounds with definite pitch that appear on both inspiratory and expiratory phase depending on the obstruction's location. They are produced when air flows through narrowed airways due to secretion or an obstructive lesion, causing the walls to vibrate. Wheezes can be further classified into two types depending on the number of airways obstructed: (1) Polyphonic wheeze and (2) Monophonic wheeze. Polyphonic wheezes are the most common type of wheezing and produce various pitched sounds. Monophonic wheezes, on the other hand, have constant pitch, timing, and site, and are caused when a single larger airway is narrowed.

Wheezes are easily recognizable as they stand out from the noise of the normal lung sound. Medium to loud intensities are especially noticeable because of the presence of sharp peaks in the power spectrum density. On the other hand, pitch is measured by the dominant frequency of the wheeze (Meslier, Charbonneau, & Racineux, 1995). Wheezes normally last up to 250 ms, but are usually longer than 80ms (Reichert et al., 2008; Pasterkamp, Kraman, & Wodicka, 1997). Their frequency range extends from 100 Hz to 1 KHz with a dominant frequency of 400 Hz (Reichert et al., 2008; Sahgal, 2011; Pasterkamp et al., 1997; Karnath & Boyars, 2002).

Ronchi

Rhonchi generate continuous, musical sounds similar to wheezes but with a coarser and lower-pitched sound. It is usually caused by air flowing through narrowed bronchial or larger airways due to secretion. This type of lung sound usually implies obstructive lung disease such as pneumonia and cystic fibrosis.

This type of lung sound contains rapidly damping periodic waveform with a duration of more than 100 ms, frequency below 300 Hz and a dominant frequency of 200 Hz (Reichert et al., 2008; Sahgal, 2011; Meslier et al., 1995). This type of breath sound is found on patients with secretions or narrowing of airways.

System Design

The acquisition device used for this work is the Thinklabs Rhythm DS32A+ Electronic Stethoscope with the volume amplification set to moderate. The electronic stethoscope has the capability to provide the sound of a conventional stethoscope without the losses of air tubing via its acoustic mode. Furthermore, it also has a diaphragm mode and bell mode used for pulmonary (higher frequency) and cardiac (lower frequency) sounds acquisition, respectively. The acoustic mode and the diaphragm mode were both enabled in this work. To consider the condition when the data acquisition device does not have noise filtering, the noise rejection capability of the device was deactivated. This was done to minimize the dependency of this work to a specific device manufacturer's performance in terms of signal processing. Lung sounds extraction was performed on the six posterior chest areas shown in Figure 1.

These raw sounds were processed and identified using the system described in Figure 2. The identification system for the electronic stethoscope is composed of four main modules: (1) the Signal Enhancement module, (2) the Feature Extraction module, (3) the Breathing Pattern Modelling module and, (4) the Breathing Pattern Classification module. The last two modules were realized using the Support Vector Machine (SVM). The following sub-sections will describe these modules further.

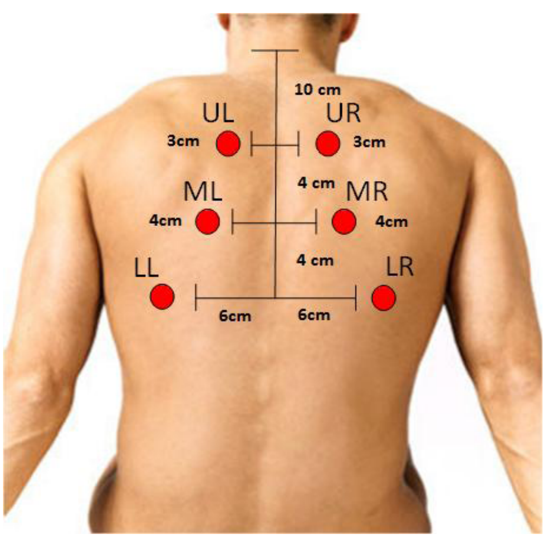

Figure 1. Six posterior chest location for auscultation

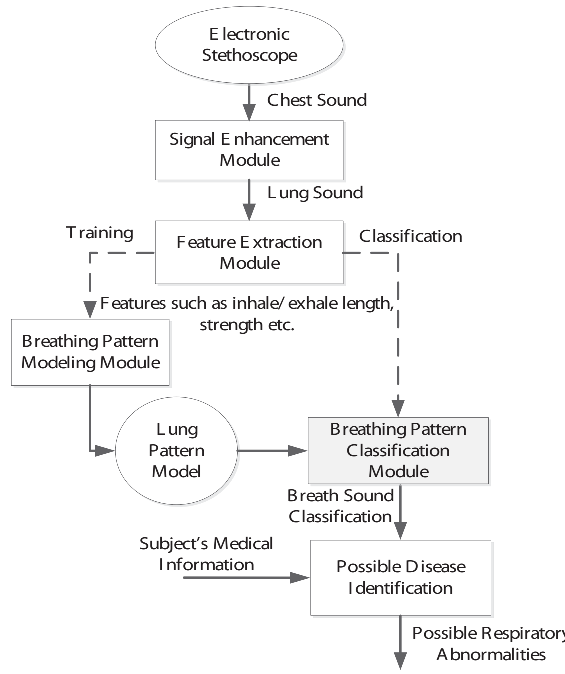

Figure 2. System Block Diagram

The Signal Enhancement module

The Signal Enhancement module performs signal denoising in order to remove the mechanical noise[1] and other out-of-lung sound band signals. Characterization of the mechanical sound present in 53 normal lung sounds was first done with the help of physician-research consultants for mechanical noise tagging.

The empirical evaluation[2] shows that the average frequency of mechanical noise is 87.34 Hz with standard deviation of 28.87 Hz. The observed range was from 31.32 Hz to 203.87 Hz, which is within the lung sound frequency range (see Table 1) such that the direct application of a notch filter on the lung sound is not possible. Further observation shows that these unwanted signals are bursty (4 to 32 ms), appearing once or twice in one breathing cycle (approximately 3 seconds). Thus, the signal denoising scheme was designed to only apply in parts of the lung sound signal where the mechanical noise is present or selective-in-time notch filter.

Figure 3 describes this joint-time frequency filtering scheme. The method applies the notch filter by first computing the Short Time Energy (STE) of the raw signal with window size equal to 40 ms corresponding to the longest mechanical noise burst, 32 ms, plus buffer.

To determine the noise-impaired portion of the raw signal, both the raw signal and the STE signal are divided into 80-ms segments, without overlap, of length equal to 40 ms. Since the mechanical noise has relatively high peak amplitude, the average power of the segment will be expectedly higher than the previous and next segments. Thus, each segment of the STE is compared with the average of the previous and next segments. If 50% of the STE peak energy exceeds the average of the previous and the next segments' peaks, then the segment being tested is said to be impaired by the mechanical noise and notch filter is applied.

The segment window size is empirically determined, evaluated at 50 ms, 60 ms, 70 ms, 80 ms, 90 ms, and 100 ms; the segment size that results in highest percentage of removed mechanical noise is used.

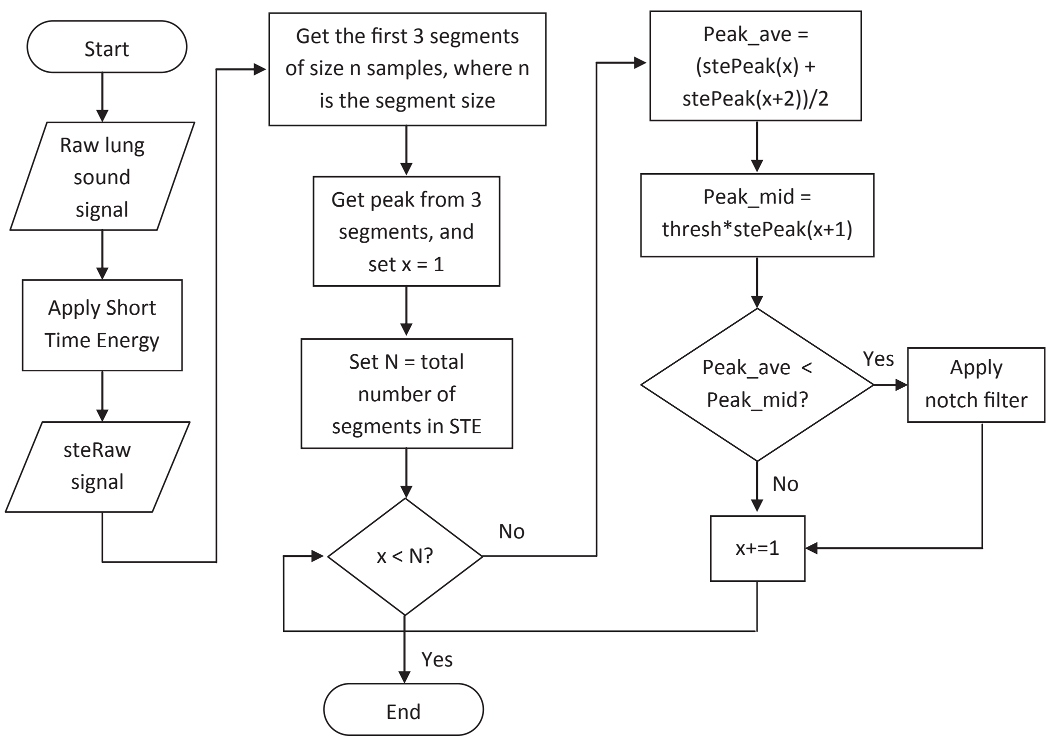

Figure 3. Joint Time-Frequency Denoising Flowchart for removing the mechanical noise in lung sounds

The Feature Extraction module

There were two groups of lung sound features that were considered in this system: (1) the Power Spectrum-based features and (2) the MFC coefficients. The lung sound characteristics show difference in terms of the pitch and intensity which could be measured in the power spectrum of the signal. Thus, the bandwidth B, center frequency f_C, and peak frequency f_{pk} , given in equations (1)-(3) respectively, were used as features.

$$B = f_H - f_L \qquad \text{Eq. (1)}$$

$$f_C = \sqrt{f_L f_H} \qquad \text{Eq. (2)}$$

$$f_{pk} = \max(\text{PSD}) \qquad \text{Eq. (3)}$$

where f_L denotes the upper cut-off frequency from DC where 5% of the total power resides, that is, , f_H denotes the upper cut-off frequency from DC where 95% of the total power resides, that is, and PSD denotes the power spectral density of the lung sound. A very low number of features may not be enough to achieve the objectives of this work that is why Mel-Frequency Cepstral Coefficients (MFCC) was also considered.

Mel-Frequency analysis is based on human perception experiments which shows that human ear acts as a filter concentrated on certain frequency components. The outputs of these filters are represented by the MFC coefficients which are typically used in speech synthesis and state-of-the-art speech recognition systems. Since lung sounds can also be analyzed in both time and frequency domains, MFC coefficients were also used as additional features for lung sound classification in this work.

MFCC is a representation of the short-time power spectrum of a sound. Mapping to Mel scale frequency is performed by dividing the signals into frames, taking the Fourier transform of each frame and then passing these to a triangular filter. Due to the perceived logarithmic characteristic of signal loudness, taking the log powers of the Mel frequencies and performing the cosine transformation will result in the MFC coefficients (MFCC). For this work, a window size equal to 15 ms with no overlap was used from 100 Hz to 2500 Hz with 15 coefficients.

The Breathing Pattern Modelling module

Support Vector Machine (SVM) is based on a statistical learning theory that classifies data by forming a hyper plane that will maximize the margin between the data sets (Lee, 2012). It has been used in several automatic classification systems such as in the studies of Phatiwuttipat, Kongprawechon, Tungpimolrut, and Yuenyong (2011) and Greenwood and Kinghorn (2012) for audio signals, achieving as high as 96.4% For this system, SVM was chosen to build the model for the lung sounds. Using the Sequential Minimal Optimization (SMO) (Platt, 1998) implementation in WEKA (Hall et al., 2009) kernel selection and parameter optimization were performed to achieve high true positive (TP)—normal lung sound is classified as normal lung sound—and low false positive (FP) —an abnormal lung sound is classified as a normal lung sound.

In the evaluation of the created model, these five metrics are considered: (1) TP, (2) FP, (3) true negative (TN), (4) false negative (FN), and (5) accuracy. Feature selection was also performed to determine the appropriate features that would separate lung sounds. Lastly, 10 fold-cross validation is used to check for overfitting, especially for this work which has a large set of parameters and relatively small amount of data.

The Breathing Pattern Classification module

This SVM-based module is the actual classification part of the system, which utilizes the SVM-based model created for the lung sounds. During the model evaluation, test data were supplied into this module to simulate the actual performance of the system.

SYSTEM TESTS, RESULTS AND ANALYSIS

The data used for the following test were gathered from the Out Patient Department (OPD) patients of the Lung Center of the Philippines and volunteers from DLSU with age ranging from 18-60 years old. Each recording is manually taken from the DS32a+ electronic stethoscope and recorded into a laptop. The software used for the recording is the Thinklabs Phonocardiography Software provided by the Thinklabs Company. For each recording, about 6 to 8 breath cycles were taken per location. These were then truncated into breathing cycles (one complete inhale and exhale phase) and separated into several wav files. The files are labeled according to its lung sound type, location in the posterior chest area, and the breath cycle count from the recording.

Each lung sound signal is equal to one breath cycle, which is composed of one inhale and one exhale phase. In summary, there are 210 normal lung sound cycles, 45 crackle sounds, 29 ronchi sounds, and 47 wheeze sounds collected. Although there is a huge amount of normal lung sound, only 29 breathing cycles per lung sound were used to balance the training and test data.

Signal Enhancement module characterization

The major features of the developed system are based on the frequency characteristics of the lung sounds. This approach was used because each type of lung sound has its own frequency range or its frequency bandwidth. For example, normal lung sounds are typically in the range of 50 Hz to 2500 Hz while ronchi are generally below 300 Hz. However, due to the inevitable movement of the stethoscope and the presence of environmental noise during auscultation, generally called here as mechanical noise, the acquired lung sounds may fall beyond its typical frequency range. These mechanical noises from data auscultation were manually characterized. In this regard, the Signal Enhancement module was introduced.

Figures 4 to 7 show the improvement in the f_L and f_H of the normal lung sounds, wheeze, crackles, and ronchi sounds, respectively. In these figures, the x-axis shows the breathing cycle samples while the y-axis indicates the frequency of the f_L and f_H. Twenty lung sound cycles for each lung sound class where used to construct the plot. For example, in Figure 4, the f_L and f_H of the normal lung sound breathing cycle number 6 is around 10 Hz and 190 Hz, respectively. After performing denoising, the f_L and f_H becomes around 250 Hz and 650 Hz.

The dotted lines show the ideal frequency range based from (Meslier et al., 1995; Sahgal 2011; Reichert 2008; Pasterkamp et al.; 1997; Karnath and Boyars, 2002). For normal lung sounds the range is from 200 Hz to 1000 Hz, for the wheeze sound the range is from 200 Hz to 2500 Hz, for crackle sounds the range is from 200 Hz to 1000 Hz, and finally for ronchi, the range is less than 300 Hz. These figures show that the raw lung sounds were drowned by mechanical noise. After denoising, its expected frequency response (f_L and f_H) were achieved.

Sequential Minimal Optimization (SMO) Kernel Selection

Two types of kernels were considered in this work: the Polynomial kernel with parameters complexity *c* and exponent *e*, and the Radial Basis Function (RBF) with parameters complexity *c* and gamma *g*. The evaluation for the kernel and the parameters to be used is organized such that the Polykernel and RBF are compared for the two models: (1) normal vs adventitious and (2) normal vs wheeze vs crackle vs ronchi. For each model, the parameters exponent and gamma were varied for Polynomial and RBF kernels, respectively. The performance metrics that were looked at are the False Positive (FP) and True Positive (TP) (Positive case being the normal lung sound) with more weight given on FP since adventitious sound should not be classified as a normal lung sound.

Table 2 provides the performance of the two kernels in classifying normal lung sounds and adventitious lung sounds from models created using Polynomial and RBF kernels. In terms of FP, both kernels have 0.0 FP rates, which is desirable. In terms of TP, the Polynomial kernel achieved the highest rate when $e = 7$.

Table 3 shows similar result from a similar test except that the model created was for specific lung sounds — normal, wheeze, crackle and ronchi. The RBF kernel-based model resulted in the lowest FP value; however, its TP rate is very low. The next desirable FP rate is at 0.028 which is the same for all kernels and parameter values except for Polynomial kernel with $e = 8$, which has 0.111 FP rate. Among the cases when FP rate = 0.028, the Polynomial kernel with $e = 7$ has the highest TP rate. The Polynomial kernel with $e = 8$ has the highest TP rate among all other FP rates. However, since higher priority is given to a lower FP rate, and considering the results in Table 2, the Polynomial kernel with $e = 7$ was chosen.

Table 2. Normal vs Adventitious Lung Sound

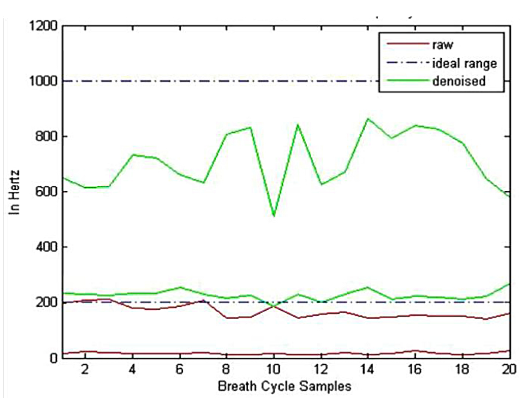

Figure 4. Lung sound frequency improvement of normal lung sounds

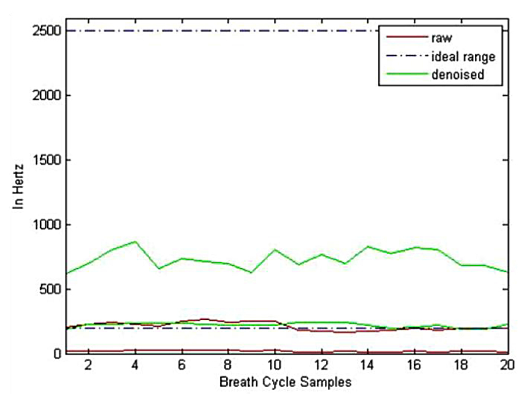

Figure 5. Lung sound frequency improvement of wheeze

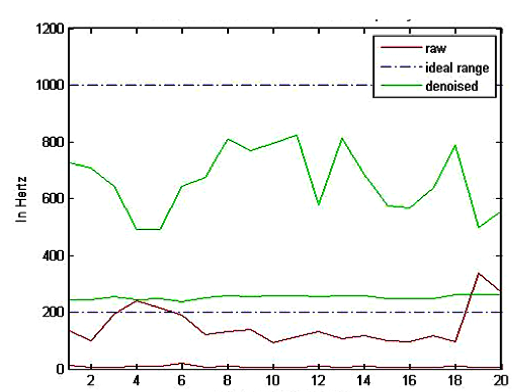

Figure 6. Lung sound frequency improvement of crackles

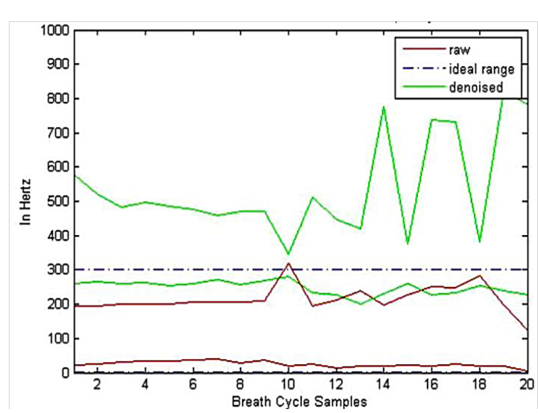

Figure 7. Lung sound frequency improvement of ronchi

Table 2. Normal vs Adventitious Lung Sound Case

Polynomial Kernel, c = 1			RBF Kernel, c = 1.5		
Exponent (e)	**TP Rate**	**FP Rate**	**Gamma (g)**	**TP Rate**	**FP Rate**
1	0.500	0.028	0.010	0.417	0.028
4	0.667	0.028	0.020	0.417	0.028
7	0.667	0.028	0.030	0. 083	0.000
8	0.917	0.111			

Table 3. Normal vs Wheeze vs Crackle vs Ronchi Case

Polynomial Kernel, c = 1			RBF Kernel, c = 1.5		
Exponent (e)	**TP Rate**	**FP Rate**	**Gamma (g)**	**TP Rate**	**FP Rate**
1	0.417	0.000	0.010	0.417	0.000
4	0.583	0.000	0.020	0.167	0.000
7	1.000	0.000	0.030	0.000	0.000
8	0.833	0.000			

Table 4. SVM-based Model using Polynomial Kernel (c = 1, e = 7) 10-Fold Cross Validation
*N = normal, A = adventitious, W = wheeze, C = crackle, R = ronch

Fold	N vs A	N vs W vs C vs R	N vs C	N vs R	N vs W
1	100.00	85.71	100.00	100.00	100.00
2	100.00	85.71	100.00	100.00	100.00
3	75.00	100.00	100.00	100.00	100.00
4	75.00	57.00	100.00	100.00	75.00
5	66.66	85.71	100.00	100.00	67.00
6	100.00	100.00	100.00	100.00	100.00
7	100.00	100.00	100.00	100.00	67.00
8	100.00	85.71	100.00	100.00	100.00
9	66.66	100.00	100.00	100.00	100.00
10	88.00	100.00	100.00	100.00	100.00
Ave	**88.00%**	**90.00%**	**100.00%**	**100.00%**	**91.00%**

Table 5. Comparison of the Correct Matches (CM) and False Matches (FM) across different Features Sets

Features sets	CM	FM
PSD-based	39.58%	60.42%
MFCC	60.42%	39.58%
MFCC + PSD-based	54.17%	45.83%
Ranked features	50.00%	50.00%

Table 6. Comparison of the True Positive (TP) and False Positive (FP) across different Features Sets

Features sets	TP	FP
PSD-based	25.00%	19.44%
MFCC	91.67%	2.77%
MFCC + PSD-based	66.67%	2.77%
Ranked features	75.00%	2.77%

Feature Selection

Feature selection is performed to optimize the learning and classification process. The features that were considered are the PSD-based features set, MFCC features set and features ranked via *Information Gain Attribute Evaluator* algorithm. The number of correct matches (CM) and false matches (FM) were compared for the three features sets using the SMO classifier with polynomial kernel, $c = 1$ and $e = 7$ to determine the best features set.

The PSD-based features are composed of the *peak frequency, center frequency, upper cutoff frequency* and *lower frequency*. Appended to this set are the MFCCs or the Mel frequency cepstral coefficients features set which is composed of 12 coefficients extracted from the lung sound framed every 80ms and 50% overlap. This resulted in 3000 features for every lung sound instance.

Table 5 and 6 show the feature selection results performed in this work. In terms of CM and FM parameters, the MFCC features set provides the best performance equal to 60.42% and 39.58%, respectively. This desirable performance is also reflected in its TP and FP rates equal to 91.67% and 2.77%, respectively, shown in Table 6. This means that PSD-based features do not contain enough information to discriminate lung sounds. Also, selecting the significant features based on the *Information Gain Attribute Evaluator* algorithm (677 features were selected out of 3000 features) does not provide good performance with CM = 50% only and TP rate = 75% only. Thus, only the MFCC coefficients are used in this system.

Cross validation

To determine the performance consistency of the model on a given data, 10-fold cross validation was performed. The results in Table 4 show the mean accuracy of the score for each fold. Five models were evaluated to determine its performance with respect to specific abnormalities namely (1) normal vs. all the adventitious lung sounds, (2) normal vs wheeze vs. crackles, vs. ronchi., (3) normal vs. crackle, (4) normal vs ronchi and (5) normal vs wheeze. For the first classification, the accuracy varies from 66.66% to 100% which is a symptom of overfitting in the training phase. This could also be seen in the normal vs wheeze vs crackle vs ronchi case and normal vs wheeze case. The result of overfitting might be attributed to the low data set collected is investigated in the following section.

Model Analysis with expanded data using SMOTE

Due to the limited number of ronchi samples, this work uses only 29 samples for each lung sound despite that there are 210 normal lung sound samples collected. This very small amount of data for training and evaluation could have affected the overfitting which is reflected in the cross validation performance of the system in Table 4. To project the system behavior at higher data samples, expanded data set was performed using the Synthetic Minority Over-Sampling Technique (SMOTE) (Hui, Wen-Yuan, & Bing-Huan, 2005). The SMOTE oversamples the minority class which are the crackle, wheeze and rhonchi to create an expanded data set with a balanced learning environment

Table 7 shows the original test confusion matrix of the balance data set using MFCC features. The original training data set has 17 lung sound cycles for each case and 12 lung sound cycles for the test set for a total of 29 lung sound cycles in the data set. Table 8 to Table 10 show the results when the data set size for each case is increased to 58 (increased by 100%), 87 (increased by 200%) and 116 (increased by 300%) lung sound cycles, respectively. For each increase in the data set, 60% of the total data was used as training set while the remaining 40% was used as the test set.

As seen in Table 7, the rhonchi and wheeze sounds are frequently misclassified having an accuracy of 16.67% and 33.33%, respectively. If the data set is increased by 100%, the accuracy in classifying the ronchi and wheeze sounds also increases to 95.65% and 60.87%, respectively. In addition, normal lung sound classification accuracy also increases to 95.65% while crackle sound classification accuracy remains to be 100%.

However, increasing the data set size further to 87 instances decreases the ronchi and normal lung sound classification accuracy to 91.43% as shown in Table 9. Conversely, the accuracy in classifying the wheeze increases to 91.43%. Furthermore, increasing the data set to 116 instances increases the accuracy in classifying normal lung sounds and wheeze to 93.33% and 100%, respectively, as shown in Table 10. However, the rhonchi identification accuracy drops to 91.11% from 91.43%. For all data expansion via SMOTE, crackle sound maintains the 100% classification accuracy.

Table 7. Confusion Matrix for data set size = 29 instances

	Classified as				
Lung sound	normal	Crackle	rhonchi	**wheeze**	**accuracy**
normal	11	0	0	1	91.67%
crackle	0	12	0	0	100%
rhonchi	1	9	2	0	16.67%
wheeze	0	5	3	4	33.33%

Table 8. Confusion Matrix for data set size = 58 instances

	Classified as				
Lung sound	**normal**	**Crackle**	**rhonchi**	**wheeze**	**accuracy**
normal	22	0	0	1	95.65%
crackle	0	23	0	0	100%
rhonchi	1	0	22	0	95.65%
wheeze	1	5	3	14	60.87%

Table 9. Confusion Matrix for data set size = 87 instances

	Classified as				
Lung sound	**normal**	**Crackle**	**rhonchi**	**wheeze**	**accuracy**
normal	32	0	0	3	91.43%
crackle	0	35	0	0	100%
rhonchi	3	0	32	0	91.43%
wheeze	1	2	0	32	91.43%

Table 10. Confusion Matrix for data set size = 116 instances

	Classified as				
Lung sound	**normal**	**Crackle**	**rhonchi**	**wheeze**	**accuracy**
normal	42	3	0	1	93.33%
crackle	0	46	0	0	100%
rhonchi	4	1	41	0	91.11%
wheeze	0	0	0	46	100%

Table 11. SVM-based Model for data set = 116 using Polynomial Kernel (c = 1, e = 7) Cross Validation

Fold	**N vs A**	**N vs W vs C vs R**	**N vs C**	**N vs R**	**N vs W**
1	100	97.78	100	100	100
2	100	97.78	100	100	100
3	100	96.67	100	100	100
4	100	98.33	100	100	97.78
5	97.78	96.67	100	100	97.78
6	97.78	97.78	100	100	100
7	100	96.67	100	100	100
8	100	96.67	100	100	100
9	100	97.78	100	100	100
10	97.78	96.67	100	100	100
Ave.	99.33%	97.28%	100.00%	100.00%	99.56%

*N = normal, A = adventitious, W = wheeze, C = crackle, R = ronchi

The increase in the number of instances also shows an improved cross validation as shown in Table 11. Classifying normal lung sound against adventitious sounds gives consistent values from 97.78%-100.00% as compared to 66.66%-100.00% in Table 4 when the data set is equal to 29 instances. Similarly, when the cross validation is performed on classifying normal lung sounds, wheeze, ronchi, and crackles, from 57.00%-100.00% across 10 folds, the accuracy becomes 96.67%-100.00% when the instances for each lung sound are increased to 116. In classifying the normal lung sound against wheeze sound, the cross validation when data set size is equal to 119 also improves from 67.00% - 100.00% to 97.78% - 100.00%.

CONCLUSION

The goals of this work are to (1) Collect chest sounds from healthy people and patients with common respiratory diseases using commercially available electronic stethoscope; (2) Determine filtering scheme that could separate enough important information of lung sounds from other chest sounds, for example, heart beat sound and external noise; (3) Manually sort and label the lung sounds into normal, wheeze, and crackle sound, with the help of physicians. This will be used as training set for the feature extraction and machine learning algorithms; (4) Determine feature extraction and machine learning algorithm that would best classify the lung sound using the patterns, and (5) evaluate the performance of the system by comparing the actual results with the lung sound.

In general, using the SMO implementation in WEKA, using Polynomial kernel with $c = 1$ and $e = 7$, there is 100% chance of identifying a normal lung sound from adventitious lung sound, that is, normal vs. adventitious. However, when tested against the three other lung sounds—normal vs. wheeze vs. rhonchi vs. crackle—the system has difficulty in differentiating wheezes and rhonchi lung sounds. Specifically, 75% of rhonchi sounds are classified as crackles and 41.67% of wheezes are classified as crackles.

Summary

Two hundred ten normal sounds, 45 crackle sounds, 29 rhonchi sounds, and 47 wheeze sounds were used in this work. These sounds were gathered from various patients in Lung Center of the Philippines and from volunteers in De La Salle University-Manila. These data were used to create a model for lung pattern classification as well as for testing. To train and test data, the lung sounds were manually verified by the medical experts from the Lung Center of the Philippines. Before recording an adventitious sound, the doctor in charge first checks the type of lung sound. The data taken are the area from the posterior chest area (Upper Left, Upper Right, Middle Left, Middle Right, Lower Left, and Lower Right), the gender, age and type of lung sound.

Processing was done using the Signal Enhancement module and the Feature Extraction module. The denoising submodule removes mechanical noise from the raw signal through selective filtering. The Feature Extraction module takes attributes from each lung sound which was used for creating a model from its pattern. Furthermore, classification and analysis are done on automatic identification on lung sound signals.

The feature extraction module uses two types of data, the PSD–based features and MFCC-based features. Features were tested and analyzed based on its confusion matrices. From these, the correct matches (CM) and false matches (FM) were computed and summarized. PSD-based features especially, which got a 39.58% CM and 25% true positive (TP) as provided in Tables 5 and 6, need to have the features extracted re-examined.

In contrast, the MFCC features, which are typically applied on speech processing, was proved to be a relatively better feature set for lung sounds with 60.42% CM and 91.67% TP, as shown in Tables 5 and 6.

Each normal lung sound was tested against adventitious lung sounds to determine the best kernel and parameters for the SMO classifier. Using the TP rate and the FP rate, the polynomial kernel showed better performance as compared to the RBF kernel with TP rate of 95% and 68.3%, respectively. The exponent $e = 7$ also gave the best FP rate equal to 1.9%. A lower FP rate indicates that less adventitious lung sounds are classified as normal lung sounds.

Despite this performance, the 10-fold cross validation shows that the system overfits or is highly dependent on the training data (please refer to Table 4). Specifically, classifying normal lung sounds against adventitious lung sounds provides varying accuracy from 66.66% to 100%, classifying specific lung sounds shows 57% to 100% accuracy, and classifying normal lung sounds against wheeze sounds gives 67% to 100%. Nevertheless, after increasing the number of instances using synthetic examples, this undesirable cross-validation result betters to 97.78% to 100% for normal vs. adventitious sounds, 96.67% to 97.78% for classifying specific lung sounds and 97.78% to 100% for normal lung sound vs. wheeze sound. This work determined that in order to address the overfitting problem of the developed system, additional training data are needed.

Contributions

The contributions of this work are as follows:

1. This work was able to collect lung sound signals consisting of 210 normal lung sounds, 47 wheezes, 45 crackles, and 29 rhonchi lung sounds. This collection may be used for lung sound-related e-Health researches, for example, signal enhancement, feature extraction, and classification.
2. This work was able to characterize the mechanical noise present with lung sound acquired using the electronic stethoscope. This noise, found to be in the range of 31-203.96 Hz, was removed using a joint time-frequency notch filter. Future work can focus on improving the noise modeling and removal process, for example, using adaptive filtering techniques on signals simultaneously acquired using two or more stethoscopes.
3. This work determined that the MFCC features are better than the usual PSD features, such as bandwidth, peak frequency and center frequency, in terms of the correct matches (CM) and sensitivity or true positive rate (TP) of the classifier. MFCC-based features can be used to classify normal vs. all with a CM rate of 60.4%, with most of the error coming from classifying the wheeze and rhonchi. In order to improve classification with other lung sounds, exploring other features, for example, Linear Predictive Coefficients, as well as improving the denoising scheme is suggested.
4. A scheme for automatic classification of lung sounds is initially investigated and developed. Table 12 shows the performance of the developed system when compared with various existing works. It should be noted that the system's reported classification accuracy of 54.16% (for normal vs. crackles vs. wheezes vs. ronchi) needs to be further improved to make it more qualified for actual deployment to rural community health services.

Table 12. Performance comparison with works and the developed system in terms of accuracy

Comparison category	Other works	Performance of this work
normal vs. adventitious	84.10% (Himeshima, Yamashita, Matsunaga, & Miyahara, 2012) 90.50% (Matsutake, Yamashita, & Matsunaga, 2013) 100.00% (Abbasi, Derakhshanfar, Abbasi, & Sarbaz, 2013) 91.71% (Kaya & Elhilali, 2013)	100.00%
normal vs. crackles	100.00% (Li & Du, 2005) 97.50% (Li, Wu, & Du, 2012)	100.00%
normal vs. wheeze	84.82% (Riella et al., 2009)	79.16%
normal vs. rhonchi	N/A[3]	83.33%
normal vs. crackle vs. wheeze vs. Rhonchi	65.37% (Emmanouilidou et al., 2012)	54.16%

Recommendations

The performance of the developed system needs improvement specially in classifying different lung sounds (normal vs. wheeze vs. crackle vs. ronchi). Increasing the number of training and test instances via the SMOTE algorithm revealed that the most probable cause of its poor performance and overfitting is the low number of data samples. It is recommended, therefore, to perform more intensive data collection with the supervision of medical experts.

This work does not consider the addition of a high pass filter that can remove artefacts of the joint time-frequency filtering. Different signal amplification factors prior to feature extraction may be explored to check how the system performance can be improved. Finally, the 2995 features used for classification may still be reduced via Principal Component Analysis or Linear Discriminant Analysis.

ACKNOWLEDGEMENT

This work was supported by the University Research Coordinating Office of DLSU. We would also like to acknowledge the physicians who provided us consultation on the medical aspect of this work: Dr. Adrian Paul J. Rabe of the Philippine General Hospital and Dr. Thendrex H. Estrella of the Rochester General Hospital.

REFERENCES

Abbasi, S., Derakhshanfar, R., Abbasi, A., & Sarbaz, Y. (2013). Classification of normal and abnormal lung sounds using neural network and support vector machines. *Electrical Engineering (ICEE), 2013 21st Iranian Conference on*, (pp. 1-4). Mashhad.

Bai, Y., & Lu, C. (2005). The embedded digital stethoscope uses the adaptive noise cancellation filter and the Type I Chebyshev IIR bandpass filter to reduce the noise of the heart sound. In *Proceedings from the 7th International Workshop on Enterprise Networking and Computing in Healthcare Industry* (pp 278-281). IEEE.

Chen, X., Han, Y., & Wu, J. (2010). Burst noise measuring on the basis of wavelet and Fourier transform. In *Proceedings from the 2010 International Conference on Measuring Technology and Mechatronics Automation* (pp

769-771). Changsha City: IEEE.

Department of Health. (2013a, April 26). *Philippine health statistics, 1960-2005.* Retrieved from http://www.doh.gov.ph/kp/statistics/selected_death

Department of Health. (2013b, April 26). *Philippine health statistics, 2001-2005.* Retrieved from http://www.doh.gov.ph/kp/statistics/morbidity#2005

Duque, F. I. (2009).A tribute to world-class pinoy doctors, Retrieved from: http://www.doh.gov.ph/system/files/messagepain.pdf..

Emmanouilidou, D., Patil, K., West, J., & Elhilali, M. (2012). A multiresolution analysis for detection of abnormal lung sounds. In *Proceedings from the 34th Annual Internation Conference of the IEEE EMBS*, (pp. 3139-3142). San Diego, California USA: IEEE.

Greenwood, M., & Kinghorn, A. (2012, July). SUVing: Automatic Silence/Unvoiced/Voiced classification of speech. Unpublished manuscript, Department of Computer Science, University of Sheffield.

Guntupalli, K. K., Alapat, P. M., Bandi, V. D., & Kushnir, I. (2008). Validation of automatic wheeze detection in patients with obstructed airways and in health subjects. *The Journal of asthma official journal of the Association for the Care of Asthma, 45*(10), 903-907.

Hall, M, Frank, E., Holmes, G., Pfahringer B., Reutemann P., Witten I. (2009). The WEKA Data Mining Software: An Update. SIGKDD Explorations, *11(1)*..

Himeshima, M., Yamashita, M., Matsunaga, S., & Miyahara, S. (2012). Detection of abnormal lung sounds taking into account duration distribution for adventitious sounds. In *Proceedings from the 20th European Signal Processing Conference (EUSIPCO 2012)*, (pp. 1821-1825). Bucharest, Romania.

Hui, H., Wen-Yuan, W., & Bing-Huan, M. (2005). Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. In H. De-Shuang, Z. Xiao-Ping, & H. Guang-Bin (Eds.), *Advances in intelligent computing lecture notes in Computer Science* (pp. 878-887). Berlin Heidelberg: Springer-Verlag.

Kahya, Y. P., Yeginer, M., & Bilgic, B. (2006). Classifying respiratory sounds with different feature sets. *Conference Proceedings in the International Conference of IEEE Engineering in Medicine and Biology Society*, (pp. 2856-2859). New York: IEEE.

Kandaswamy, A., Kumar, C. S., Ramanathan, R. P., Jayaraman, S., & Malmurugan, N. (2004). Neural classification of lung sounds using wavelet coefficients. *Computers in Biology and Medicine, 34*(6), 523-537.

Karnath, B., & Boyars, M. (2002, January). Pulmonary Auscultation. (B. Karnath, Ed.) *Hospital Physicians: Review of clinical Signs*, pp. 22-26. Retrieved from http://www.turner-white.com/rcs/jan02/rcs_jan02_pulm.htm.

Kaya, E., & Elhilali, M. (2013). Abnormality detection in noisy biosignals. In *Proceedings from the 35th Annual International Conference of the IEEE EMBS*, (pp. 3949-3952). Osaka, Japan: IEEE.

Lee, Y. (2012). Introduction to support vector machine and their applications in bankcruptcy prognosis. In Y.-J. Lee, Y.-R. Yeh, & H.-K. Pao (Eds.), *Handbook of computational finance* (pp. 731-761). Berlin Heidelberg: Springer.

Li, Z., & Du, M. (2005). HHT-based lung sound crackle detection and classification. In *Proceedings from the 2005 International Symposium on Intelligent Signal Processing and Communication Systems*, (pp. 385-388). Hong Kong: IEEE.

Li, Z., Wu, X., & Du, M. (2012). A novel method for feature extration of crackles in lung sound. In *Proceedings from the 5th International Conferen on BioMedical Engineering and Informatics (BMEI 2012)*, (pp. 399-402).

Lu, X., & Bahoura, M. (2008). An integrated automated system for crackles extraction and classification. *Biomedical Signal Processing and Control, 3*(3), 244-254.

Matsutake, S., Yamashita, M., & Matsunaga, S. (2013). Discrimination between healthy subjects and patients using lung sounds from multiple auscultation points. In *Proceedings from the Acoustic, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on* , (pp. 1296-1300). Vancouver, BC: IEEE.

Meslier, N., Charbonneau, G., & Racineux, J. (1995). Wheezes: Series Chest Physical Examination. *United Kingdom: European*

Respiratory Journal 8(11), 1942-1948.

Pasterkamp, H., Kraman, S. S., & Wodicka, G. R. (1997, September). Respiratory sounds. *American Journal of Respiratory and Critical Care Medicine, 156*(3), 974-987.

Phatiwuttipat, P., Kongprawechon, W., Tungpimolrut, K., & Yuenyong, S. (2011). Cardiac auscultation analysis system with neural network and SVM techniques. In *Proceedings from the Electrical Engineering/ Electronics, Computer, Teecommunications and Information Technology (ECTI-CON), 2011 8th International Conference on*, (pp. 1021-1030). Khon Kaen: IEEE.

Piirila, P., & Sovijarvi, A. (1995). Crackles recording, analysis and clinical significance: Series chest physical examination. *European Respiratory Journal 8*(12), 2139-2148.

Platt, J. (1998). *Sequential minimal optimization: A fast algorithm for training support vector machines*. Technical Report MSR-TR-98-14, Microsoft Research.

Reichert, S., Gass, R., Brandt, C., & Andres, E. (2008, May). Analysis of respiratory sounds: State of the art. *Clinical medicine: circulatory, respiratory and pulmonary medicine, 2*, 45-58.

Riella, R. J., Nohama, P., & Maia, J. M. (2009). Method for automatic detection of wheezing in lung sounds. *Brazilian journal of medical and biological research, 42*(7), 674-684.

Sahgal, N. (2011). Monitoring and analysis of lung sounds remotely. *International journal of chronic obstructive pulmonary disease, 6*, (pp. 407-412).

Sakai, T., Kato, M., Miyahara, S., & Kiyasu, S. (2012). Robust detection of adventitious lung sounds in electronic auscultation signals. In *Proceedings from the 21st International Conference on Pattern Recognition (ICPR 2012)*, (pp. 1993-1996). Tsukuba, Japan. IEEE.

Samaniego, T. S. (2011, September 30). Helping improve quality of healthcare across Southeast Asia. *Philippine Daily Inquirer*. http://business.inquirer.net/22277/helping-improve-quality-of-healthcare-across-southeast-asia

Schmidt, S., Holst-Hansen, C., Graff, C., Toft, E., & Stuijk, J. (2007). Detection of coronary artery disease with an electronic stethoscope. In *Proceedings from the IEEE Conference on Computers in Cardiology*.

Shah, G., & Papadias, C. (2013). Separation of cardiorespiratory sounds using time-frequency masking and sparsity. In *Proceedings from the Digital Signal Processing (DSP), 2013 18th International Conferencen on.* Fira: IEEE.

Varady, P. (23-25 October 2001). Wavelet-based adaptive denoising of phonocardiographic records. In *Proceedings from the 23rd Annual EMBS International Conference*. Istanbul, Turkey: IEEE.

Waitmann, L. R., Clarkson, K. P., Barwise, J. A., & King, P. H. (2000). Representation and classification of breath sounds recorded in an intensive care setting using neural network. *Journal of Clinical Monitoring and Computing, 16*(2), 95-105.

Yu, F., Bilberg, A., & Voss, F. (2008). The development of an intelligent electronic stethoscope. In *Proceedings from the IEEE/ ASME International Conference on Mechatronic and Embedded Systems and Applications*. Beijing: IEEE.