Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: A systematic review and meta-analysis

Arati Gurunga, Carolyn G. Scrafford, James M. Tielsch, Orin S. Levine, William Checkley

Division of Pulmonary and Critical Care, School of Medicine, Johns Hopkins University, Baltimore, MD, USA
Program in Global Disease Epidemiology and Control, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

Received 11 March 2011; accepted 11 May 2011
Available online 14 June 2011

KEYWORDS
Pneumonia; Respiratory disorders; Electronic auscultation; Lung sound analysis

Summary
Rationale: The standardized use of a stethoscope for chest auscultation in clinical research is limited by its inherent inter-listener variability. Electronic auscultation and automated classification of recorded lung sounds may help prevent some of these shortcomings.
Objective: We sought to perform a systematic review and meta-analysis of studies implementing computerized lung sound analysis (CLSA) to aid in the detection of abnormal lung sounds for specific respiratory disorders.
Methods: We searched for articles on CLSA in MEDLINE, EMBASE, Cochrane Library and ISI Web of Knowledge through July 31, 2010. Following qualitative review, we conducted a meta-analysis to estimate the sensitivity and specificity of CLSA for the detection of abnormal lung sounds.
Measurements and main results: Of 208 articles identified, we selected eight studies for review. Most studies employed either electret microphones or piezoelectric sensors for auscultation, and Fourier Transform and Neural Network algorithms for analysis and automated classification of lung sounds. Overall sensitivity for the detection of wheezes or crackles using CLSA was 80% (95% CI 72–86%) and specificity was 85% (95% CI 78–91%).
Conclusions: While quality data on CLSA are relatively limited, analysis of existing information suggests that CLSA can provide a relatively high specificity for detecting abnormal lung sounds such as crackles and wheezes. Further research and product development could promote the value of CLSA in research studies or its diagnostic utility in clinical settings.

© 2011 Elsevier Ltd. All rights reserved.
Introduction

The stethoscope is used by clinicians to aid in the diagnosis of respiratory disorders; however, application of the stethoscope in research studies has been limited due to the inherent inter-observer variability and subjectivity in the interpretation of lung sounds. The diagnostic value of auscultation in detecting abnormal lung sounds in clinical research could be improved if implemented using an objective and standardized approach to interpretation. Computerized analysis of recorded lung sounds may offer a systematic approach to the diagnosis of different respiratory conditions via automated classification of acoustic patterns.

Computerized lung sound analysis involves recording the patient’s lung sounds via an electronic device, followed by computer analysis and classification of lung sounds based on specific signal characteristics. Auscultation typically takes place in a clinic setting where there could be multiple sources of ambient noise. Acoustic auscultation, however, is generally limited by poor signal transmission due to noise, tubular resonance effects, and greater attenuation of higher frequency sounds. This is an important factor to consider in pulmonary auscultation because lung sounds are mostly characterized in the higher frequency spectrum ranging from 50 Hz to 2500 Hz. On the other hand, electronic auscultation has the advantage of signal amplification and ambient noise reduction leading to increased signal-to-noise ratio along with its independence on human ear sensitivity to different acoustic frequencies.

Rapid advancement in electronics and computer technology in recent years has increased research interest in automated classification of lung sounds among pulmonary researchers and has the potential to reduce software and hardware costs. Computerized lung sound analysis is a powerful tool for optimizing and quantifying electronic auscultation information based on the specific lung sound spectral characteristics. The Fourier transform is the most commonly used spectral analysis algorithm to provide information on the frequency components of a signal. Neural network, a machine learning algorithm for feature recognition and classification, has been used for classification of different lung sounds based on the features selected from frequency decomposition and associated statistical parameters.

We sought to perform a systematic review on computerized lung sound analysis and investigate its diagnostic utility in classifying recorded auscultatory findings. Our primary objective was to estimate the overall sensitivity and specificity of computerized lung sound analysis algorithms for detection of abnormal lung sounds based on currently available studies.

Methods

Systematic review methods

Two investigators (AG, CGS) conducted independent literature searches using MEDLINE and EMBASE until July 31, 2010. Additional studies were obtained from references of identified studies, the Cochrane Library and the ISI Web of Knowledge. We searched for studies relating to use of chest auscultation and computer algorithms for automated detection and classification of lung sounds. Keywords included “pneumonia”, “acute lower respiratory infections (ALRI)”, “lung auscultation”, “electronic auscultation”, “Acoustic signal processing”, “computerized lung sound analysis”, “automated classification of lung sounds”, “crackle detection”, “wheeze detection” and “World Health Organization guidelines”. Literature search was based on corresponding Medical Subject Headings (MeSH) suggested in the MEDLINE database. We restricted our search to English because this was the only language for which both investigators were fluent. We excluded case reports or manuscripts not based on original research.

The two reviewers independently reviewed and compared the resulting list of relevant articles and determined the eligibility of the full report. We evaluated each of the included studies for the method of electronic auscultation and the analytical approach used for the classification of lung sounds. We rated the quality of the studies based on the Newcastle–Ottawa Quality Assessment Scale. We followed the Meta-analysis of Observational Studies in Epidemiology guidelines.

Biostatistical methods

Both reviewers individually extracted data from selected studies and compared outcomes for validation. Data from each selected study were summarized in a contingency table that compared CLSA performance against the gold standard diagnosis (i.e., chest radiography or clinical diagnosis). Four studies were selected for qualitative review only because they did not provide sufficient data to construct a contingency table or did not employ comparable methodology for statistical computation of sensitivity and specificity. E.g. used number of crackles rather than number of cases and controls. We estimated overall sensitivity and specificity from the four studies selected for meta-analysis using a fixed-effects model. We assessed statistical heterogeneity between studies using Cochrane Q-tests and I²-values. We calculated a summary receiver operating characteristic (ROC) curve using standard techniques. We used R (www.r-project.org) for statistical analysis.

Results

Overview of the literature search

We identified 208 articles via our initial search. After reading the abstract and methodology, we selected 40 articles for further review. Based on a detailed review of these 40 studies, we selected a total of 12 studies that had relevant methodology for qualitative discussion. Duplicate articles were discarded which led to identification of eight studies for systematic review out of which only four studies had sufficient quantitative information for a meta-analysis (Table 1).
<table>
<thead>
<tr>
<th>Study (year)</th>
<th>Location</th>
<th>N</th>
<th>Age in years as Mean ± SD or range</th>
<th>Gender Ratio (M:F)</th>
<th>Methodology</th>
<th>Case definition</th>
<th>Recording Device</th>
<th>Algorithm</th>
<th>Chest sound evaluated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rietveld et al. (1999)</td>
<td>Amsterdam, Netherlands</td>
<td>Cases: 50, Controls: 10</td>
<td>7–18 (mean = 12.1, SD = 2.9)</td>
<td>N/A</td>
<td>Asthma cases were recruited upon physicians’ referral whereas controls were enrolled via advertisements on newspapers in the Netherlands. Two experienced examiners conducted data analysis blinded to patient information. Case severity was verified by British lung physicians.</td>
<td>Asthma</td>
<td>Microphones</td>
<td>Fourier Transform Neural Network</td>
<td>Wheezing</td>
</tr>
<tr>
<td>Waitman et al. (2000)</td>
<td>Nashville, Tennessee</td>
<td>Cases and controls: 17</td>
<td>19–75</td>
<td>11:6</td>
<td>Lung sounds were collected from patients admitted in the Neurocritical Intensive Care Unit at Vanderbilt University Hospital and audio tapes used to train their clinicians. Two physicians conducted data analysis in a quiet environment blinded to patient information.</td>
<td>Pathological lung sounds</td>
<td>Microphones</td>
<td>Fourier Transform Neural Network</td>
<td>Crackles, wheezes, pneumonia, pleural rub, squawks, stridor</td>
</tr>
<tr>
<td>Murphy et al. (2004)</td>
<td>Boston, Massachusetts</td>
<td>Cases: 100, Controls: 100 (50% learning; 50% test)</td>
<td>69 ± 18</td>
<td>Cases: 42:58, Controls: 52:48</td>
<td>Subjects were recruited from the community teaching hospital in Boston, MA. Cases were confirmed by radiographic findings that were evaluated by two blinded, independent observers. Cases and controls were matched based on their age.</td>
<td>Pneumonia confirmed with radiographic evidence</td>
<td>Microphones</td>
<td>Time Expanded Waveform Analysis, Acoustic Power Analysis</td>
<td>Crackles, rhonchus, wheeze — used to generate an “acoustic pneumonia score”</td>
</tr>
<tr>
<td>Study</td>
<td>Location</td>
<td>Cases</td>
<td>Controls</td>
<td>Samples</td>
<td>Subjects</td>
<td>Diagnosis</td>
<td>Pneumonia or pleural effusion</td>
<td>Piezo-electric sensors</td>
<td>Fourier Transform, Vibration Response Imaging (VRI)</td>
</tr>
<tr>
<td>---------------------</td>
<td>------------------</td>
<td>-------</td>
<td>----------</td>
<td>---------</td>
<td>----------</td>
<td>-----------</td>
<td>-----------------------------</td>
<td>------------------------</td>
<td>---------------------------------------------------</td>
</tr>
<tr>
<td>Mor et al. (2007)</td>
<td>Israel</td>
<td>20</td>
<td>60</td>
<td>N/A</td>
<td>N/A</td>
<td>Radiologically and clinically confirmed diagnosis of pneumonia or pleural effusion</td>
<td>56 ± 17 (20 learning; 40 test)</td>
<td>Piezo-electric sensors</td>
<td>Fourier Transform, Vibration Response Imaging (VRI)</td>
</tr>
<tr>
<td>Guntupalli et al. (2008)</td>
<td>Houston, Texas</td>
<td>7</td>
<td>7</td>
<td>28–75</td>
<td>N/A</td>
<td>Asthma and COPD</td>
<td>Piezo-electric sensors</td>
<td>Crackles</td>
<td>Wavelet Transform k-Nearest Neighbor</td>
</tr>
<tr>
<td>Ono et al. (2009)</td>
<td>Tokyo, Japan</td>
<td>21</td>
<td>10</td>
<td>65.7 ± 11.0</td>
<td>Similar in the two groups</td>
<td>Interstitial pneumonia</td>
<td>Microphones and Transducer</td>
<td>Crackles</td>
<td>Wavelet Transform k-Nearest Neighbor</td>
</tr>
<tr>
<td>Riella et al. (2009)</td>
<td>Curitiba, Brazil</td>
<td>NA</td>
<td>NA</td>
<td>0–76</td>
<td>N/A</td>
<td>Lung sound recordings were selected from internet repositories for validation of methodology</td>
<td>NA</td>
<td>Short-time Fourier Transform, Neural Network</td>
<td>Crackles</td>
</tr>
</tbody>
</table>
Characteristics of selected studies

All but one of the selected studies was conducted among adult populations from a clinical setting, either from an intensive care unit or a community hospital. Rietveld et al. recruited children and adolescents diagnosed with asthma. All of the studies were from middle- and high-income countries. Guntupalli et al. recruited subjects with asthma. Murphy et al., Mor et al., and Ono et al. recruited subjects with pneumonia. Waitman et al. and Kahya et al. selected patients with abnormal lung sounds upon auscultation.

Three studies specifically conducted computerized lung sound analysis for classification of wheezing,\(^9\)–\(^{11}\) two analyzed crackles\(^{12,14}\) and three analyzed both wheezes and crackles.\(^1,10,13\) Murphy et al. included a customized pneumonia score to scale the diagnosis of pneumonia which included the number of crackles identified in the respiratory cycle. Kahya et al. also included a customized crackles parameter, based on the duration of crackles during a respiratory cycle. Both studies indicated that inclusion of additional parameters like the number of crackles improved sensitivity and specificity of the computerized lung sound algorithm in classifying different lung sounds. Kahya et al. found an increase in sensitivity and specificity from 80% to 95% and 80% to 90%, respectively during the inspiratory phase, and found an improvement in specificity from 85% to 95% during the expiratory phase.

We also sought to describe the types of recording devices and automated detection and classification algorithms employed by the eight studies. Five studies used microphones for electronic auscultation\(^1,8,10,12,14\) and two used piezoelectric sensors.\(^13,17\) Riella et al. used lung sounds available electronically from different online repositories and indicated the common usage of a contact accelerometer for recording though the authors did not provide information on the recording setting. Most of the studies selected one or four breath cycles segmented into inspiratory and expiratory phases with average recording duration ranging from 3 to 20 s. Mor et al. was the only study that specified instructing the patients to take deeper breaths than normal during recording which took place in a quiet but not sound proof room. Ono et al. was the only study that specified recording lung sounds in a sound proof room.

Five studies employed Fourier transform algorithms for lung sound classification (Table 1). One study used derivatives of Fourier Transform namely the Short-Time Fourier Transform and Wavelets. One included the Neural Network algorithm. One employed a k-Nearest Neighbor algorithm whereas two others classified lung sounds based on a two-dimensional gray-scale imaging system called vibration response imaging.

Quality of selected studies

Based on the Newcastle–Ottawa Quality Assessment Scale, Murphy et al. and Mor et al. each scored an average of 7.5, Ono et al. scored an average of 6.5, Guntupalli et al. scored an average of 4.5, whereas Rietveld et al., Waitman et al. and Kahya et al. each scored an average of 4 (Table 2). Riella et al. did not meet any of the Newcastle–Ottawa Quality Assessment Scale criteria. While all of the selected studies, except for Riella et al., provided a detailed description on inclusion criteria for their cases and applied the same study design for both the cases and the controls, only two studies provided information on selection criteria for controls\(^1,14\) and none required post-intervention responses from the subjects.

Meta-analysis

All four studies included for meta-analysis provided more than one dataset based on the data collection methodology. Ono et al. reported results from two independent observers, B and D. We randomly selected data from Observer B to be included in the meta-analysis. A sensitivity analysis using data from Observer D yielded similar results (data not shown). Mor et al. also conducted two separate analyses, one where the physician was blinded to patient information and one with patient information provided. We decided to use the blinded analysis due to the potential for diagnosis bias. Murphy et al. provided a separate statistical dataset for a learning sample and a test sample for use in the Neural Network algorithm employed for automated classification. We decided to use the data from the test sample in our analysis. Kahya et al. provided results for various parameters used in the analysis. Subgroup statistical analysis in the Kahya et al. study included data during the inspiratory versus the expiratory phases. Based on the significance of these two respiratory phases in the pathology of pneumonia, we decided to include inspiratory and expiratory separately in our meta-analysis.

In Table 3, we display the data we obtained from our review of the four studies included in the meta-analysis. Sensitivity and specificity for the selected studies varied from 70% to 95% and from 80% to 95%, respectively. The CLSA algorithm had an overall sensitivity of 80% (95% CI 72–86%, Fig. 1) and an overall specificity of 85% (95% CI 78–91%, Fig. 2). We display the ROC curve for studies included in the meta-analysis in Fig. 3. Q-values for overall sensitivity and specificity were 4.9 (\(p = 0.30\)) and 3.1 (\(p = 0.55\)), respectively; \(I^2\)-values were 18% (95% CI 0–83%) and 0% (95% CI 0–73%), respectively.

Discussion

Computerized analysis of recorded lung sounds may be a promising adjunct to chest auscultation as a diagnostic aid in both clinical and research settings. In our meta-analysis, we found that computerized lung sound analysis performs at a relatively high level of sensitivity and specificity in a small number of studies. Overall sensitivity for the detection of abnormal lung sounds using computerized lung sound analysis was 80% and overall specificity was 85%.

Our systematic review revealed, however, that there is a lack of standardization across studies in the methods used for lung sound recording, computer algorithms for signal analysis and statistical methods for outcome analysis. For example, recording lung sounds in a noisy clinic versus a quiet
research room would demand a more rigorous post-
processing technique to combat the noise present in the
acoustic signal and the efficiency of the classification algo-


der would vary accordingly. Such inconsistency not only

Table 2  Quality Assessment Score for Selected Studies Based on the Newcastle-Ottawa Quality Assessment Scale by Reviewer.

<table>
<thead>
<tr>
<th>Study</th>
<th>Selection</th>
<th>Comparability</th>
<th>Exposure</th>
</tr>
</thead>
</table>
| [Average Score] | Adequate Case Definition | Represen-
| tativness of the cases | Selection of controls | Definition of controls | Controls for age | Controls for gender | Ascertainment of Exposure | Same method for cases and controls | Non-response rate |
| Rietveld 1999 [4] | * | * | * | * | * | * | * | * |
| Waitman 2000 [4] | * | * | * | * | * | * | * | * |
| Murphy 2000 [7.5] | * | * | * | * | * | * | * | * |
| Mor 2007 [7.5] | * | * | * | * | * | * | * | * |
| Kahya 2008 [4] | * | * | * | * | * | * | * | * |
| Guntupalli 2008 [4.5] | * | * | * | * | * | * | * | * |
| Ono 2009 [6.5] | * | * | * | * | * | * | * | * |
| Rietta 2009 [0] | * | * | * | * | * | * | * | * |

The number of stars assigned indicates the quality of the study based on the Newcastle-Ottawa Scale. For example, Murphy et al. included two items specified under the Ascertainment of Exposure category and was assigned two stars for higher quality.

Figure 1  Sensitivity Plot for Studies Included in Meta-

Figure 2  Specificity Plot for Studies Included in Meta-

The Sensitivity of individual studies are represented

The Specificity of individual studies are represented

by squares and corresponding 95% confidence intervals are

represented by segments. The overall Sensitivity and corre-

sponding 95% confidence intervals represented by a diamond.

The center of the diamond is the overall Sensitivity and the

length of the diamond is the 95% confidence interval. E =

Expiratory, I = Inspiratory.
leads to difficulty in interpreting and translating study outcomes but also has prevented commercialization of computerized lung sound analysis devices. Further research is needed to address the effectiveness of specific combinations of electronic devices and computing algorithms in clinical and community settings. In the advent of electronic auscultation and advanced signal processing techniques, Andres et al. has initiated a collaborative project, "Analyse des Sons Auscultatoires Pathologiques" (ASAP), a French national program that aimed to develop a database of objective definitions of auscultation sounds characteristic of certain pathologies, standardized formats of lung sound recordings and storage to facilitate exchange of information among health care providers. Likewise, Sovijarvi et al. has developed a set of Computerized Respiratory Sound Analysis (CORSA) guidelines to standardize the definitions and terminologies used in computerized lung sound analysis. Further evaluation and advancement of ASAP and CORSA guidelines may help standardize computerized lung sound analysis techniques and definitions and promote its research and development.

The selection of an appropriate signal processing technique is important in improving the diagnostic quality of the algorithm. A growing number of studies have analyzed the efficacy and efficiency of spectral analysis algorithms in detecting and classifying wheezes and crackles. Lung sounds range in frequency between 50 Hz and 2500 Hz, and tracheal sounds can reach values up to 4000 Hz. Normal breath sounds have their main frequency band between 200 Hz and 250 Hz, and normal tracheal sounds range from 850 Hz to 1000 Hz. Abnormal breath sounds have higher and wider frequency band; continuous wheezes lie between 100 Hz and 2500 Hz with a dominant frequency between 100 Hz and 1000 Hz, and have a duration greater than 100 ms. Crackles are in the frequency range of 100 Hz to 2000 Hz or even higher, with two cycle duration of greater than 10 ms for coarse crackles and two cycle duration of less than 10 ms for fine crackles. Therefore, frequency analysis and quantification of biological signals such as lung sounds provides information that is not readily available in the time-domain and thereby allows such signal processing algorithms to automatically distinguish between normal and abnormal lung sounds. Inclusion of a broader spectrum of detection parameters, from both the time and the frequency domains might further improve the sensitivity and specificity of the computerized lung sound analysis algorithm in detecting pneumonia and other respiratory disorders. Spectral analysis with inclusion of lung sound related features such as crackle counts and duration, specific inspiration versus expiration, was found to improve the diagnostic accuracy.

Our study is limited by the small number of available studies on the clinical utility of auscultation and computerized lung sound analysis for the diagnosis of abnormal lung sounds. Selected studies widely involved adult populations with the exception of Rietveld et al. who included children and adolescents with a mean age of 12 years. Gross et al. indicated that the spectral components of lung sounds have slight variance among different age groups, particularly in infants and children. Therefore, the findings of this study could not be directly applied to a pediatric population.

In summary, electronic auscultation coupled with computerized lung sound analysis has the potential to improve diagnostic yield of pulmonary disorders in both clinical and research settings. More studies are needed to characterize recorded lung sounds, particularly in children. Furthermore, our study identified the need to address standardization of methodology and analytical methods across studies. This technology, however, faces a significant challenge before achieving clinical acceptance due to

![Figure 3](image-url)  
**Figure 3** Summary Receiver Operating Characteristics Plot for CLSA algorithms. Individual studies are displayed as circles. Overall sensitivity and specificity is displayed as a cross.

<table>
<thead>
<tr>
<th>Study</th>
<th>TP (%)</th>
<th>FP (%)</th>
<th>FN (%)</th>
<th>TN (%)</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murphy et al.</td>
<td>39 (39%)</td>
<td>6 (6%)</td>
<td>11 (11%)</td>
<td>44 (44%)</td>
<td>100</td>
</tr>
<tr>
<td>Mor et al.</td>
<td>14 (23%)</td>
<td>8 (13%)</td>
<td>6 (10%)</td>
<td>32 (53%)</td>
<td>60</td>
</tr>
<tr>
<td>Kahya et al. (Expiratory)</td>
<td>18 (45%)</td>
<td>1 (3%)</td>
<td>2 (5%)</td>
<td>19 (47%)</td>
<td>40</td>
</tr>
<tr>
<td>Kahya et al. (Inspiratory)</td>
<td>19 (47%)</td>
<td>2 (5%)</td>
<td>1 (3%)</td>
<td>18 (45%)</td>
<td>40</td>
</tr>
<tr>
<td>Ono et al.*</td>
<td>17 (55%)</td>
<td>2 (6%)</td>
<td>4 (13%)</td>
<td>8 (26%)</td>
<td>31</td>
</tr>
</tbody>
</table>

TP = True Positives.  
FP = False Positives.  
FN = False Negatives.  
TN = True Negatives.  
* Classification based on data from Observer B.
Meta-analysis of computerized lung sound analysis algorithms


References