Annals of Biomedical Engineering (© 2013)
DOI: 10.1007/s10439-013-0836-0

BMES BIOMEDICAL ENGINEERING SOCIETY™ www.bmes.org

Cough Sound Analysis Can Rapidly Diagnose Childhood Pneumonia

UDANTHA R. ABEYRATNE,[1] VINAYAK SWARNKAR,[1] AMALIA SETYATI,[2] and RINA TRIASIH[2]

[1]School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD, Australia; and [2]Respiratory Medicine Unit, Sardjito Hospital, Gadjah Mada University, Yogyakarta, Indonesia

(*Received 27 January 2013; accepted 27 May 2013*)

Associate Editor Merryn Tawhai oversaw the review of this article.

Abstract—Pneumonia annually kills over 1,800,000 children throughout the world. The vast majority of these deaths occur in resource poor regions such as the sub-Saharan Africa and remote Asia. Prompt diagnosis and proper treatment are essential to prevent these unnecessary deaths. The reliable diagnosis of childhood pneumonia in remote regions is fraught with difficulties arising from the lack of field-deployable imaging and laboratory facilities as well as the scarcity of trained community healthcare workers. In this paper, we present a pioneering class of technology addressing both of these problems. Our approach is centred on the automated analysis of cough and respiratory sounds, collected *via* microphones that do not require physical contact with subjects. Cough is a cardinal symptom of pneumonia but the current clinical routines used in remote settings do not make use of coughs beyond *noting its existence* as a screening-in criterion. We hypothesized that cough carries vital information to diagnose pneumonia, and developed mathematical features and a pattern classifier system suited for the task. We collected cough sounds from 91 patients suspected of acute respiratory illness such as pneumonia, bronchiolitis and asthma. Non-contact microphones kept by the patient's bedside were used for data acquisition. We extracted features such as non-Gaussianity and Mel Cepstra from cough sounds and used them to train a Logistic Regression classifier. We used the clinical diagnosis provided by the paediatric respiratory clinician as the gold standard to train and validate our classifier. The methods proposed in this paper could separate pneumonia from other diseases at a sensitivity and specificity of 94 and 75% respectively, based on parameters extracted from cough sounds alone. The inclusion of other simple measurements such as the presence of fever further increased the performance. These results show that cough sounds indeed carry critical information on the lower respiratory tract, and can be used to diagnose pneumonia. The performance of our method is far superior to those of existing WHO clinical algorithms for resource-poor regions. To the best of our knowledge, this is the first attempt in the world to diagnose pneumonia in humans using cough sound analysis. Our method has the potential to revolutionize the management of childhood pneumonia in remote regions of the world.

Keywords—Pneumonia, Childhood cough, Automated cough analysis.

INTRODUCTION

Pneumonia is the leading killer of young children around the world. It accounts for more than 19% of under-five child deaths each year.[26] The vast majority of these deaths occur in resource poor regions of the world such as sub-Saharan Africa, South Asia and remote areas of China and Indonesia. It is also a common childhood illness in Australian indigenous communities. Childhood pneumonia is largely a disease of poverty and is often called the "forgotten disease".

The *definitive* diagnosis of childhood pneumonia, especially the early stage disease, is surprisingly difficult even in a hospital. Lung biopsy may be the most effective approach but it is clearly impractical for clinical use. The clinical examination and chest auscultation with a stethoscope are the first steps in diagnosing childhood pneumonia. Chest X-ray is often used as an important reference standard in confirming a clinical diagnosis. However, X-rays may not be sensitive to early stage pneumonia[3] or when the diseased part of the lung is not clearly visible on the image. In addition, normal X-ray may lead to poor specificity of diagnosis in the presence of lung scarring or congestive heart disease. X-ray CT imaging (Computed Tomography) and other laboratory analyses such as sputum tests, blood culture and C-reactive protein (CRP) tests may be needed to differentially diagnose pneumonia in some cases.

Unfortunately, even primitive imaging and laboratory testing facilities are beyond the reach for much of

Address correspondence to Udantha R. Abeyratne, School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD, Australia. Electronic mail: udantha@itee.uq.edu.au

Published online: 07 June 2013

the global population. In resource-poor areas of the world where pneumonia is rampant, it is also difficult to find trained healthcare personnel with expert auscultation and clinical skills. The management of pneumonia in such regions is largely dependent on community workers who visit remote communities. Difficulties in the timely diagnosis and proper treatment are the main reasons behind the unacceptably high rate of childhood pneumonia deaths (1.8 million/year[26]) in the world.

In order to address these problems, the World Health Organization (WHO) has developed a simple clinical algorithm (see Table 1 for details) to classify pneumonia in resource-limited regions.[10] These classifications directly lead to interventions such as antibiotic prescription and hospitalization. The WHO algorithm uses the symptom of cough (and/or breathing difficulty) as the screening-in feature for pneumonia; breathing rate then determines if pneumonia exists. The disease will be further classified as severe pneumonia if symptoms such as chest recession and stridor are also present. In the past, several studies[4,5,8,18] have explored the performance of the WHO algorithm in pneumonia diagnosis. They reported a reasonably high sensitivity (69–94%) but an unacceptably poor specificity (16–67%).

The poor specificity means the WHO algorithm has a high false positive rate, which results in an over-prescription of antibiotics and the wastage of rare drug stocks. It also has led to antibiotic resistance[3,20] in vulnerable communities. Treatment failure rates as high as 22% have been observed in some regions.[3] Failure to respond to the limited variety of inexpensive antibiotics available in resource-limited regions has also become a serious problem in the prompt and proper treatment of childhood pneumonia. Complicating matters further, there is a symptom overlap between pneumonia and malaria, which is another disease commonly found in many pneumonia endemic regions. The low specificity of the WHO algorithms leads to misdiagnosis, causing time delays in starting proper treatment.

Two of the critical challenges to be met in managing the global burden of childhood pneumonia are the low diagnostic performance of the WHO algorithm and difficulties in training the large number of workers[24] needed for community healthcare visits. It is known that most deaths from childhood pneumonia occur early in the progression of the disease, and accurate diagnosis and prompt treatment by community health workers may reduce pneumonia mortality by 36–42%[24] in resource-poor regions. Researchers have attempted to improve the specificity of the WHO criteria using different approaches. These include the augmentation of WHO algorithm by considering fever[4] and other symptoms of pneumonia (nasal flaring, poor sleep, chest in-drawing, cough lasting longer than two days *etc.*). These efforts resulted in a sensitivity and specificity within the range 20–90%,[4,8,13,15,18] but higher specificities were achieved only at the cost of lower sensitivity and vice versa. They also suffer from the fact that the higher the complexity of measurements, the more difficult it is to train community workers to reliably implement the algorithm in field visits.

TABLE 1. WHO/IMCI guidelines for pneumonia classification in resource-poor regions.

Screening in criteria	Cough and/or breathing difficulty.
Criterion for Pneumonia	Fast breathing. The definition of fast breathing depends on the child's age: 2–11 months: ≥50 breaths per minute 12–60 months: ≥40 breaths per minute
Criterion for Severe Pneumonia	Fast breathing with Lower chest wall in-drawing, and/or Stridor in calm child

In this paper, we propose a pioneering class of technology addressing these challenges. Our methods are centred on the analysis of cough sounds in the diagnosis of pneumonia. The simplest form of the proposed technology does not use sensors that require physical connection with the patient, making it easy to use in the field. Our ultimate target is to deliver an automated device suitable for use by laypeople with minimal or no healthcare training. Our aim is to achieve a higher specificity of diagnosis than that of the WHO algorithm while maintaining a diagnostic sensitivity $>90\%$.

Cough is a cardinal symptom of pneumonia, but the mathematical analysis of cough has never been used in diagnosing the disease. In the WHO algorithm for resource-poor regions, only the *existence of cough* is used as a screening-in criterion. While cough exists in almost all pneumonia patients, its existence alone is not a specific enough marker of the disease. The vast potential in extracting more information from cough sounds remains virtually untapped in the current diagnosis of pneumonia.

In this work we hypothesize that cough carries vital information on the lower respiratory tract enabling us to identify pneumonia-specific features. Support for this hypothesis comes from the patho-physiology of pneumonia, our prior explorations,[12,23] previous studies[1,11,19,25] on cough sound analysis in patients with other respiratory diseases and the physics of cough sound generation. We are also encouraged by the fact that experienced pediatricians can use *qualitative differences* in coughs and other respiratory sounds such as stridor to aid in the diagnosis of some respiratory diseases,[6] and even glean some information on the place of origin of the sounds. However, the field

is in its infancy and "pneumonia-cough" has received no description other than "pained coughs that stop abruptly".[6]

Pneumonia is defined as an infection in the lungs with an accumulation of inflammatory cells and secretions[22] in the alveoli. We hypothesise that the inflammation and secretions alter the acoustic properties of respiratory sounds, including coughs. Throughout millennia physicians have listened to respiratory sounds directly or aided by devices such as stethoscopes. In auscultation, chest musculature filters out high frequency components bringing the bandwidth below 4 kHz. Current knowledge on cough is inherently bounded by the limitations of human perception (20 Hz–20 kHz nominally, but the upper frequency limit rarely extending beyond 15 kHz in middle-aged people). Our approach to cough analysis relies on the fact that cough sounds emanating from the mouth *via* the air column has a much higher bandwidth than traditional stethoscope based signals, which are severely low-pass filtered by the chest musculature.

We hypothesize that the disease of pneumonia, with its characteristic features such as lung consolidation, fluid-filled alveoli, compromised capacity for airflow through the affected parts of the lung and popping noise (crackles) from alveoli alters the acoustics of cough generation. These differences are further accentuated by the fact that consolidated parts of the lung conduct sound better and faster, compared to the normal lung. Our approach is to extract resulting acoustical features and develop signal processing and pattern recognition technology to diagnose pneumonia.

In this paper we describe the technology we developed to diagnose childhood pneumonia, and evaluate its performance on a clinical database of respiratory sounds.

MATERIAL AND METHOD

The overall approach we propose in this paper is summarised in Fig. 1. The method consists of three main stages: the acquisition of data from subjects, the development of features and the training of pattern classifiers, and, the prospective validation and performance evaluation of the technology. In "Cough Sound Recording Environment and Protocols", "The Development of the Cough Sound Database", "Feature Extraction and the Pattern Classifier Design", and "Performance Estimation on the Prospective Dataset D_{pv}" sections, we provide details of these steps.

Cough Sound Recording Environment and Protocols

The clinical data acquisition environment for this work is the Respiratory Medicine Unit of the Sardjito Hospital, Gadjah Mada University, Indonesia. Patients suspected of acute respiratory illnesses such as

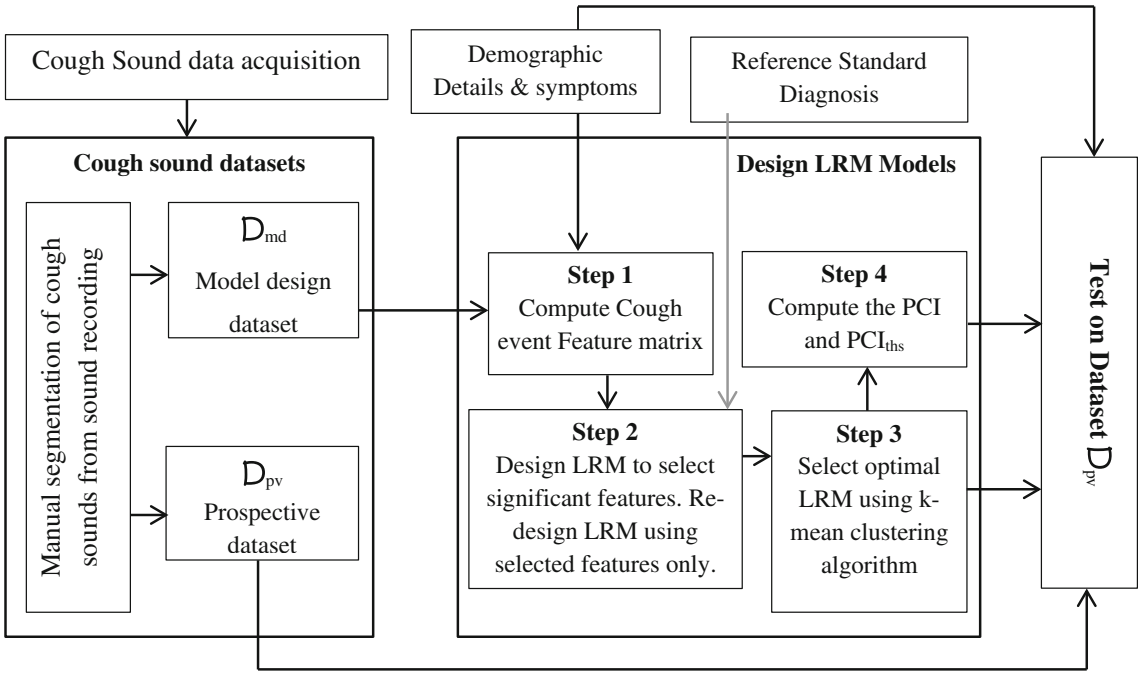

FIGURE 1. Overall block diagram of the proposed method. The method consists of three main stages: (i) creating cough sound datasets, (ii) the development of features & the training of pattern classifiers, and, (iii) the prospective validation and performance evaluation of the technology.

pneumonia, bronchiolitis, bronchitis, asthma and rhino-pharyngitis were recruited for the study. Patients were recruited within the first 12 h of their admission. Table 2 lists the inclusion and exclusion criteria of subjects. The human ethics committees of Gadjah Mada University and The University of Queensland had approved the study protocol and the patient recruitment procedures.

A paediatrician clinically assessed patients upon their presentation to the hospital. Symptoms such as the presence of cough, runny nose, fever, breathing difficulty and diarrhoea were documented using a standard form. Routine demographic information as well as the outcomes of the clinical assessment (e.g. chest auscultation, breathing rate, oxymetry, temperature) were also recorded. The reference standard used for Pneumonia diagnosis in this paper is the overall diagnosis provided by physicians, on the basis of clinical presentation, laboratory tests (e.g.: blood, sputum analysis), chest X-ray and the clinical course of the disease. In order to minimize the radiation exposure to children, X-ray was performed only on subjects clinically suspected of pneumonia, or if there was a clear clinical need for it. Thus, not all subjects in our database had undergone X-ray imaging.

After the initial medical assessment, sound recordings were made in the natural environment of the respiratory ward. Most of the recordings were made in single occupancy rooms, but occasionally patients shared the room with another patient, in spaces separated by curtains. A family member often accompanied paediatric patients.

Figure 2 shows the instruments used to record high fidelity cough sounds from patients. The system consisted of two bedside microphones (Rode® NT3, 44.1 kHz sampling rate, cardioid directional pattern). One microphone was aimed directly towards the patient and the other one in the opposite direction. This facilitated the differentiation of cough sounds of the subject from background interference. The major sources of background sound were: adult speech, footsteps, toy noise and sounds generated by trolley as well as door movements.

In this paper, our target is to develop technology robust against sound intensity variations, such that the distance from the mouth to the recording device or the transduction sensitivity of the microphone do not play any significant role in the diagnosis. In our sound recordings, the distance between the patients and the microphone aimed towards them varied between 40 and 70 cm depending on the position of their heads at a given time.

The length of sound data recording in our dataset varied from 11 min to 12 h and 34 min (average duration 4 h and 3 min with a standard deviation of 1 h and 39 min). The time window over which we could record varied due to a variety of reasons beyond our control. For instance, we had to terminate recording sessions due to: adults coming for long visitations, a child falling asleep for a long period, a busy diagnostic facility (e.g. X-ray) becoming available and a patient requiring to be immediately sent away for imaging, mains power failures, and patients becoming unavailable due to medical emergencies.

The Development of the Cough Sound Database

We used a total of 91 patients (63 pneumonia and 28 non-pneumonia subjects) to develop and validate our technology. Diseases such as bronchiolitis, asthma, bronchitis, pharyngitis, laryngomalacia are lumped within the non-pneumonia group. The overall dataset at our disposal was separated into two non-overlapping groups: the *Model Development Dataset* (D_{md})

TABLE 2. Inclusion and exclusion criteria used in the study.

Inclusion criteria	Exclusion criteria
Patients with symptoms of chest infection : at least 2 of	Advanced disease where recovery is not expected e.g. terminal lung cancer
Cough	Droplet precautions
Sputum	NIV required
Increased breathlessness	No informed-consent
Temperature >37.5°	
Consent	

FIGURE 2. Cough sound instrumentation at the hospital. Sound recordings were made in the respiratory ward of the Sardjito Hospital, Gadjah Mada University, Indonesia. Most of the recordings were made in single occupancy rooms, but occasionally patients shared the room with another patient.

and the *Prospective Validation Dataset* ($\mathbf{D}_{pv}$). These two datasets were completely independent of each other. The sets $\mathbf{D}_{md}$ and $\mathbf{D}_{pv}$ did not share either coughs or any subjects. Patients were assigned to each group based on the order of presentation to the respiratory clinic of the hospital. At the time of writing this paper $\mathbf{D}_{md}$ and $\mathbf{D}_{pv}$ consisted of $N_{md} = 66$ (46 pneumonia and 20 non-pneumonia patients) and $N_{pv} = 25$ (17 pneumonia and 8 non-pneumonia patients) subjects respectively.

There is no widely accepted method for the automatic identification of coughs. The manual analysis is still used as the gold standard[21] in clinical work as well as in research literature. For the work of this paper, we manually segmented out cough sounds after a careful listening process. For the training dataset $\mathbf{D}_{md}$, researchers involved in this work picked cough events from the continuous sound recording. Our target in this paper was to use first 5–10 clean coughs per patient to develop the model. In some cases (<15% of the cases), due to reasons such as short recordings, patient sleeping, background sound interference as described in "Cough Sound Recording Environment and Protocols", section less than 5 clean coughs were found and included in the training dataset.

In the Prospective Dataset ($\mathbf{D}_{pv}$), a researcher unrelated to this paper identified coughs. He was blind to the clinical diagnosis of the patients. His task was set as picking the first 15 events of coughs in a recording uncorrupted with audio interferences such as speech from visitors, trolley noise, sounds from toys, and doors closing. The beginning and end of the cough events were marked after careful listening while simultaneously looking at the cough waveform and its spectrogram, using the software *Adobe Audition® CS5.5*.

Feature Extraction and the Pattern Classifier Design

We used the Model Development Dataset $\mathbf{D}_{md}$ for the work described in this section. Let C_{md} be the total numbers of cough events from the subjects in $\mathbf{D}_{md}$. The approach taken in this paper uses a leave-one-out model building and validation process to develop features and optimize model parameters. Within this framework, our method can be described in four major processing steps (see Fig. 1), Step 1 to Step 4 as described below.

In Step 1 we compute a feature matrix. In Step 2 we design automatic classifiers using the feature matrix from Step 1 to classify cough sounds into 'pneumonic cough' and 'non-pneumonic cough' classes. In Step 3 we select an optimal classifier and in the fourth step we define a new index called Pneumonic Cough Index (PCI) to identify patients with pneumonia.

Step 1: Extraction and Augmentation of Cough Features

In this step, our first target is to extract features from cough sounds to be used in the pneumonia diagnosis algorithm. Mathematical features from each cough event in $\mathbf{D}_{md}$ were computed as follows (Please see Appendix for details):

(i) Let $\mathbf{x}$ denotes a discrete time sound signal from an arbitrary cough event.
(ii) Segment $\mathbf{x}$ into three equal sized non-overlapping sub-segments. The decision to divide coughs in such segments is largely mathematical. Our target is to capture the variation of mathematical features within a single cough. Let $\mathbf{x}_i$ represents the ith sub-segment of $\mathbf{x}$, where $i = 1, 2, 3$.
(iii) For each of the sub-segments $\mathbf{x}_i$ compute the following features: the Bispectrum Score (BGS), the Non-Gaussianity score (NGS), the first four formant frequencies (FF), log energy (LogE), zero crossing (ZCR), kurtosis (Kurt), and twelve Mel-frequency cepstral coefficients (MFCC).
(iv) Repeat steps (i)–(iii) for all C_{md} cough events in $\mathbf{D}_{md}$.

This process leads to a candidate cough feature matrix $\mathbf{M}_c$ of the size $C_{md} \times C_f$. Where $C_f = 63$ represents the total number of features computed from $\mathbf{x}_i$ and C_{md} is the total coughs in the data set $\mathbf{D}_{md}$.

In the simplest form of the diagnostic algorithm, we will only be using cough-based features to diagnose pneumonia. Our fundamental focus in this work is to develop technology to diagnose pneumonia in resource-poor remote regions of the world. The ultimate target is to fully automate the technology and implement them on low-cost devices for use by community outreach workers. Thus, both the simplicity and the high diagnostic performance are essential features of the approach. Furthermore, cough sounds can be acquired with sensors that do not require physical contact with patients.

However, we recognize the existence of some simple clinical measurements that may be used to improve our algorithms at a minimal extra cost and complexity. The existing WHO algorithm for resource-poor areas (Table 1) uses the age and breathing rates, and other researchers have attempted to use the existence of fever. While none of these alone or in combination have yielded the desired diagnostic performance in remote areas, these measurements have the potential to augment cough-derived features. Inspired by the WHO algorithm that uses age as one of the parameters, we used age in months as a candidate parameter in our models. We also used the presence or absence of fever

as a binary variable that can assume the values 'yes' or 'no'.

In the WHO algorithm, breathing rate is used as the prime parameter in diagnosing pneumonia. In our work, we propose a new measure (see (1)), which we call the Breathing Index (BrI), to capture breathing rate elevations in pneumonia.

$$\mathrm{BrI} = \left\{ \begin{array}{ll} \mathrm{BR} - 20 & \text{if Age} \geq 60 \text{ months} \\ \mathrm{BR} - 40 & \text{otherwise} \end{array} \right\} \quad (1)$$

In (1) BR is breathing rate and Age is age of the patient in months. While fever is a common symptom of pneumonia, it is not specific to pneumonia. A similar observation holds for the breathing rate. We define the candidate *Augmented Feature* set by: $\mathbf{F}_{\mathrm{ac}} = \{\boldsymbol{f}_{0\mathrm{c}}, f_1, f_2, \ldots, f_{\mathrm{f}}\}$, where $\boldsymbol{f}_{0\mathrm{c}}$ represents the cough-based candidate feature vector (set) and $\{f_1, f_2, \ldots, f_{\mathrm{f}}\}$ denotes the *Clinical Feature Set* used to augmented $\boldsymbol{f}_{0\mathrm{c}}$ in our models.

We denote the candidate feature matrix based on the Augmented Feature Set by $\mathbf{M}_{\mathrm{ac}}^{(i)}$, where the index '*i*' symbolises different subsets of features we can choose from $\{f_1, f_2, \ldots, f_{\mathrm{f}}\}$. The final features in our algorithms will be drawn from the group of candidate features. Details of the feature selection, model development and validation will be described in Step 2 below.

Step 2: Feature Selection and Automatic Classifier Design

In this paper we used a Logistic-regression model (LRM) as the pattern classifier. LRM is a generalized linear model, which uses several independent features to estimate the probability of a categorical event (dependent variable). In this work, the dependent variable Y is assumed to be equal to "one" ($Y = 1$) for pneumonic cough and "zero" ($Y = 0$) for non-pneumonic cough. Cough events drawn from a subject with a diagnostic classification of pneumonia are labelled pneumonic coughs and vice versa. A model is derived using a regression function to estimate the probability Y given the independent cough features (i.e. $\mathbf{F}_{\mathrm{ca}} = \{\boldsymbol{f}_{0\mathrm{c}}, f_1, f_2, \ldots, f_{\mathrm{f}}\}$) as follows:

$$\mathrm{Prob}\left(Y = 1|_{\boldsymbol{f}_{0c}, f_1, f_2, f_3, \ldots, f_{\mathrm{F}}}\right) = \frac{e^z}{e^z + 1} \quad (2)$$

$$z = \beta_0 + \boldsymbol{\beta}_1 \cdot \boldsymbol{f}_{0c} + \beta_2 f_1 + \cdots + \beta_{n-1} f_{\mathrm{F}} \quad (3)$$

In (3), β_0 is called the intercept and $\boldsymbol{\beta}_1$, β_2 and so on are called the regression coefficients of independent variables. Note that since $\boldsymbol{f}_{0c}$ is a vector, $\boldsymbol{\beta}_1$ is also a vector and $\boldsymbol{\beta}_1 . \boldsymbol{f}_{0c}$ denotes the inner product. To select the optimal decision threshold λ from Y (that the cough is pneumonic if $Y > \lambda$; non-pneumonic otherwise) we used the Receiver-Operating Curve (ROC) analysis.

We used a Leave-One-Out validation (LOOV) technique for the LRM design. As the name suggests, LOOV technique involves using data from all the patients except one to train the model and cough events from the remaining patient to validate the model. This process was systematically repeated such that each patient in D_{md} was used to validate a model exactly one time. At the end of this process, we end up in N_{md} different LRM models. To evaluate the performance of the designed N_{md} LRMs, performance measures such as Sensitivity, Specificity, Accuracy, Positive Predicted Value (PPV), Negative Predicted Value (NPV) and Cohen's Kappa (K) statistics were computed.

(i) *Feature Selection*: Feature selection is a technique of selecting a subset of relevant features for building a robust classifier. Optimal feature selection requires the exhaustive search of all possible subsets of features. However, it is impractical to do so for the large number of features we use as candidate features. Therefore, we used an alternative approach based on the *p* value to determine significant features. In LRM design, a *p*-value is computed for each feature and it indicates how significant that feature is to the model. Important features have low *p*-value. We used this property of LRM to select a reasonable combination of features that facilitate the classification, in the model during the training phase. Our approach consisted of computing the mean *p*-value for $\boldsymbol{f}_{0\mathrm{c}}$ features over N_{md} LRMs, and then selecting the features with mean *p*-value less than a threshold p_{ths}. Let $\boldsymbol{f}_{0\mathrm{s}}$ be the subset of selected cough-based features from the candidate set $\boldsymbol{f}_{0\mathrm{c}}$. We will denote the feature matrix computed from $\boldsymbol{f}_{0\mathrm{s}}$ by $\mathbf{M}_{\mathrm{cs}}$.

(ii) *Robust LRM design*: Once the subset $\boldsymbol{f}_{0\mathrm{s}}$ is known, we use it to build a new set of LRMs once again, following another LOOV process. At the end of this process, we have N_{md} number of LRMs using $\boldsymbol{f}_{0\mathrm{s}}$ as the input feature set. The next step is to find the best LRM out of these. We describe the procedure we followed in Step 3 below.

Step 3: Selecting a Good Model From N_{md} *LRMs*

From the candidate LRMs that use the selected features $\boldsymbol{f}_{0\mathrm{s}}$ as the input feature vector, we selected one as the *Best Model* based on the k-means clustering algorithm. In the k-mean clustering algorithm, the

target is to divide q data points in d-dimensional space into k clusters so that within a cluster the sum of squared distance from the centroid is minimized.

Our target is to select the best possible model from the collection of N_{md} models available to us. For that purpose, we divided our N_{md} models in d-dimensional space into $k = 2$ clusters, i.e. the high-performance model cluster and the low-performance model cluster. We then set the space dimension d equal to the model parameters plus three performance measures (sensitivity, specificity and kappa). Then, from the cluster of the high-performance models, we selected that model which had the lowest mean square error value with respect to the centroid. Let $\Re_{0s}$ represent the selected LRM and λ_{0s} be the corresponding probability decision threshold (value determined using ROC curves such that the classifier performance is maximized) for a specific combination of features.

Once $\Re_{0s}$ is chosen, we fix all the parameters of the model and completely terminate the training process. The model $\Re_{0s}$ is then used as the best model to classify each individual cough event into 'pneumonic-cough' or 'non-pneumonic-cough' groups.

Step 4: Pneumonic Cough Index

In this step, for each $\mathsf{N}_{md} = 66$ patient in the D_{md} we computed a Pneumonic Cough Index (PCI) using the definition given in (4).

The Definition of Pneumonic Cough Index (PCI)

Let P be the total number of coughs analysed from a patient and Q is the number of coughs that were classified as pneumonic using selected LRM $\Re_{0s}$ in Step 3. Then the PCI index for a patient is computed as the fraction of a patient's coughs that were classified as pneumonic.

$$\mathrm{PCI} = Q/P \tag{4}$$

Then using the ROC analysis we computed a threshold PCI_{th} (optimized for high sensitivity while keeping acceptable specificity) to classify patient into two classes, 'Pneumonia' and 'non-Pneumonia'.

Performance Estimation on the Prospective Dataset D_{pv}

In this section, our target is to explore the performance of the method on the prospective dataset D_{pv}. For this, we fixed the model parameters and all decision thresholds at their best values as determined in "Feature Extraction and the Pattern Classifier Design". Since the sets D_{md} and D_{pv} are mutually exclusive both in patients and cough events, performance measures we compute over D_{pv} are independent of the model training process.

Using cough events from $\mathsf{N}_{pv} = 25$ patients in the dataset D_{pv}, we computed the feature matrices and then used the model $\{\mathbf{M}_{cs}, \Re_{0s}\}$ to classify each cough into the classes 'pneumonic' and 'non-pneumonic'. Then we used (4) to compute the PCI index for each patient in D_{pv}. Patients were then classified as pneumonia subjects if $\mathrm{PCI} > \mathrm{PCI}_{th}$. We then compared the results of our automatic classifications with the reference method described in "Cough Sound Recording Environment and Protocols" section, and computed performance statistics.

RESULTS

The Subject Database and the Clinical Diagnosis

For the work of this paper, we used sound recordings from 91 patients (48 males and 43 females). The mean age of the subjects was 3 years and 1 month (standard deviation 3 years and 11 months). The age of the subjects varied from 1 month to 15 years (3 neonatal subjects; 36 subjects between 2 and 12 months; 32 subjects between 13 and 60 months, and, 20 subjects aged >60 months). Of the 91 subjects, 63 were diagnosed with pneumonia and 28 were non-pneumonia patients. Non-pneumonia patients had diseases such as Asthma (8 patients), Bronchitis (8 patients), Rhino pharyngitis (5 patients) and others (2 heart disease, 1 each of wheezing, bronchiolitis, tonsilopharyngitis, foreign body inhalation and kidney failure patient). Chest X ray (CXR) was performed on 65 patients to confirm the diagnosis. Of the 26 patients on whom CXR was not done, eight had been clinically diagnosed as pneumonic and 18 as non-pneumonia patients.

Characteristics of Cough Sounds

In Fig. 3 we show typical cough samples (and their spectrograms) from patients diagnosed with pneumonia, asthma, bronchitis and bronchiolitis. We can see differences among cough groups supporting the hypothesis that cough carries information on the state of the lower respiratory tract. For instance, the pneumonia-cough tends to be of shorter duration with the power spectrum extending up-to 20 kHz. The mean duration of pneumonia-cough (0.26 ± 0.07 s using $n = 306$ coughs, training dataset D_{md}) was less (2-tailed t test, $\rho < 0.005$, $t = -5.06$ than that of other cough (0.30 ± 0.08 s using $n = 134$ coughs, training dataset D_{md}).

The cough sound waveforms were generally clean with a high signal-to-noise-ratio (SNR). The mean SNR for the D_{md} was 15.8 ± 5.6 dB (maximum = 28.05 dB and minimum = 2.08 dB) and that for D_{pv} was 16.7 ± 5 dB (maximum = 26.7 dB and minimum = 7.9 dB).

Pneumonia Diagnosis Based on WHO Criteria for Resource-Poor Regions

Table 3 shows the contingency table for pneumonia diagnosis using the WHO criteria vs. clinically diagnosed pneumonia. This table allows us to compare the performance of the WHO algorithm for resource-poor regions against our reference standard. Since the WHO guidelines are designed for children within the age group of 2–60 months,[24], in generating Table 3 we excluded subjects outside that age group. In total, 68 subjects in our database were found to be within the age range of 2–60 months. The WHO criteria achieved a high sensitivity of 83% in picking clinically confirmed pneumonia cases, but resulted in a poor specificity of 47%. Our target is to develop technology to exceed this performance.

Performance of the Proposed Method on the Model Development Dataset D_{md}

As described in "The Development of the Cough Sound Database", we divided our total of 91 subjects into two mutually exclusive datasets; the Model Development Dataset (D_{md}) and the Prospective Validation Dataset (D_{pv}). In this section we illustrate the characteristics of the proposed method and show its performance on D_{md}. The D_{md} had data from $\mathsf{N}_{md} = 66$ patients from which we had extracted $C_{md} = 440$ cough events (the average number of cough per subject = 6.7 ± 2).

Feature Matrix Computation

Following the method given in "Feature Extraction and the Pattern Classifier Design" section, Step 1, we

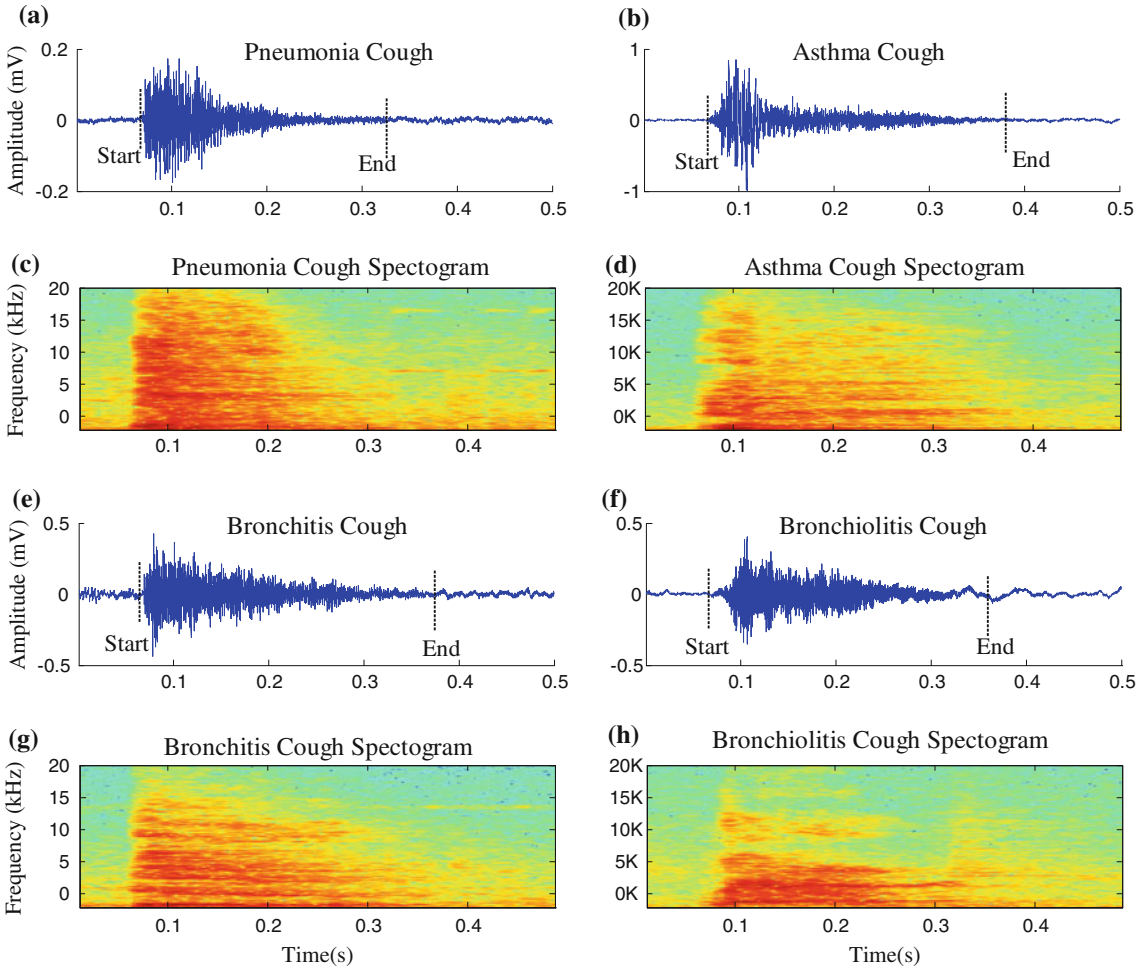

FIGURE 3. Typical waveforms of cough sounds in (a) pneumonia, (b) asthma, (e) bronchitis and (f) bronchiolitis. Their frequency spectrograms are shown respectively in (c), (d), (g) and (h).

TABLE 3. Contingency table for comparing the performance of the *WHO criteria for resource-poor regions* against our reference standard diagnosis.

		Reference Standard Diagnosis	
		Pneumonia	**Non-Pneumonia**
WHO Criteria Based Diagnosis	**Pneumonia**	44	8
	Non-Pneumonia	9	7

The WHO criteria resulted in a diagnostic sensitivity of 83%, a specificity of 47% and an accuracy of 75%. WHO criteria are only defined for children in the age group of 2-60Â months.[24] For this reason, we excluded all the subjects outside this age from our calculations. A total of 68 subjects were included in the computations.

computed the feature matrix $\mathbf{M}_c$. We divided each cough segment into 3 sub-segments for analysis, which resulted in 63 mathematical features from each cough event. These consisted of 3×12 *MFCC* coefficients, 3×4 formants, and $3\times$ (1 each of BSG, NGS, LogE, ZCR and Kurt features).

Using the feature matrix $\mathbf{M}_c$ we developed $N_{md} = 66$ Logistic Regression Models (LRM) following the Leave-One-Out Validation (LOOV) procedure described in "Feature Extraction and the Pattern Classifier Design" section (Step 1 and Step 2). Note that N_{md} is the number of subjects in the D_{md}.

Classification of Individual Cough Events Based on the Candidate Cough Feature Set

At first, we explored the performance of the cough-based candidate feature set f_{0c} in the LOOV training and validation process. The mean *training* sensitivity and specificity over the 66 LRMs in classifying individual cough events into pneumonic and non-pneumonic classes were both at $81 \pm 1\%$. The validation sensitivity and specificity were 63 and 52% respectively. Table 4a shows the detailed performance of the models on the validation dataset.

Classification of Individual Cough Events Based on Selected Cough Features

The performance of the method with the overall candidate cough-based feature set f_{0c} indicates the possibility that cough sounds carry non-trivial information on pneumonia. However, the performance requires further improvement.

Following the procedure described in "Feature Extraction and the Pattern Classifier Design" section (Step 2) and using a threshold for the p-value at $p_{th} = 0.23$, we identified the most significant features of f_{0c} to obtain the selected feature set f_{0s}. The set f_{0s} consisted of 30 cough-based features as follows: 1 feature from NGS, 2 each from BSG and LogE, 3 each from FF and ZCR and 19 MFCC features. Out of the original 63 features in f_{0c}, 33 features were dropped from the selected feature set as insignificantly contributing to the LRM model.

Table 4b gives the performance of the LRM using the feature set f_{0s}. The LOOV procedure described in "Feature Extraction and the Pattern Classifier Design" section (Step 1 and Step 2) was used to arrive at these results. With $f_{0s,}$ the sensitivity and specificity of individual cough classification have increased to 69 and 64% respectively, from their previous values at 63 and 52% with f_{0c}.

Diagnosis of Patients Based on the Pneumonia Cough Index (PCI)

So far, we explored the process of labelling each cough event in the recording as pneumonic or non-pneumonic. However, in practice, our ultimate target is to classify diseases at the patient level. In order to do that, we followed the method detailed in "Feature Extraction and the Pattern Classifier Design" section (Step 3 and Step 4). We obtained a set of $N_{md} = 66$ LRMs using the feature vector f_{0s} with the dataset D_{md} and subjected the set of LRMs to a k-means clustering procedure as described in "Feature Extraction and the Pattern Classifier Design" section (Step 3). The best LRM (denoted by $\Re_{0s}$) was chosen and its parameters were fixed. The model $\{\mathbf{M}_{cs}, \Re_{0s}\}$ was used to label each cough event as pneumonic or non-pneumonic. Then the PCI index was computed for each patient.

In order to determine the best threshold for the PCI index in diagnosing pneumonia, we did an ROC analysis, comparing the PCI-based diagnosis against our reference standard diagnosis. We found the best PCI threshold (PCI_{th}) to be 0.5. Table 5 shows the PCI based pneumonia classification results at the patient level. Using only cough-based features, we were able to obtain a sensitivity of 93% and a specificity of 90.5% on the Model Development dataset. The Cohen's Kappa for the results was given by $\kappa = 0.83$.

Diagnosis of Patients Based on the Augmented Feature Sets

The results we illustrated so far used only cough-based features as inputs to the classifier model. In this section we investigate the effect of augmenting cough-features with other simple clinical features that are currently being used in diagnosing pneumonia in resource-poor regions.

Augmenting the cough-based selected feature set $\boldsymbol{f}_{0s}$ with a combination of clinical features (Breathing Index (BrI), the existence of fever, and the age) we also generated five Augmented Feature Matrices $\mathrm{M}_{\mathrm{as}}^{(i)}$, $i = 1,2,\ldots5$. The subsets of clinical features used to generate $\mathrm{M}_{\mathrm{as}}^{(i)}$ were: (a) the BrI ($i = 1$), (b) the Existence of fever ($i = 2$), (c) the age ($i = 3$), (d) the BrI and the existence of fever ($i = 4$), and, (e) the BrI, age and the existence of fever ($i = 5$). The best performing models corresponding to the feature matrices, $\mathrm{M}_{\mathrm{as}}^{(i)}$, as determined *via* the LOOV process on the dataset D_{md}, are denoted by the symbol $\Re_{is}$, where $i = 1,2,\ldots5$.

The performance of the models $\{\mathrm{M}_{\mathrm{as},}^{(i)}\Re_{is}\}$, $i = 1,2,\ldots5$ are shown in Table 5. Note that $\{\mathbf{M}_{\mathrm{cs}}, \Re_{0s}\}$ denotes the model that uses only cough-based features. When the cough features were augmented with the BrI, i.e., in the model $\{\mathrm{M}_{\mathrm{as},}^{(i)}\Re_{is}\}$, the sensitivity and the specificity of individual cough classifications jumped respectively to 72 and 69%, in the validation data (Table 4b). The sensitivity and specificity of patient-level pneumonia diagnosis increased to 93 and 90.5% respectively.

In the literature[4] it is suggested that adding fever to the WHO algorithm improves its specificity. Inspired by this, we augmented cough-features with the binary feature 'existence of fever' in the model $\{\mathrm{M}_{\mathrm{as},}^{(2)}\Re_{2s}\}$. This model achieved sensitivities and specificities of 72 and 60% in classifying individual coughs events, and 91 and 86% in diagnosing pneumonia at the patient level. The best performance on the individual cough classification (sensitivity 80%, specificity 73%) and patient level pneumonia diagnosis (sensitivity 95.6%, specificity 90.5%) were achieved with the model $\{\mathrm{M}_{\mathrm{as},}^{(5)}\Re_{5s}\}$ where cough-features are augmented with age, BrI and the existence of fever.

Performance of the Method on the Prospective Dataset D_{pv}

Results described in "Performance of the Proposed Method on the Model Development Dataset D_{md}" section have been derived from the Model Development dataset (D_{md}) through a leave-one-out validation

TABLE 4. Performance of the 66 LRM models in classifying individual cough events into pneumonic and non-pneumonic classes.

	Sens (%)	Spec (%)	Acc (%)	PPV (%)	NPV (%)	κ
(A) Results with all the candidate cough features ($\boldsymbol{f}_{0c}$)						
Candidate Cough Features ($\boldsymbol{f}_{0c}$) only	63	52	59	73	40	0.13
$\boldsymbol{f}_{0c}$ augmented with BrI	71	55	66	77	48	0.26
$\boldsymbol{f}_{0c}$ augmented with (existence of) fever	68	54	63	75	45	0.20
$\boldsymbol{f}_{0c}$ augmented with Age	70	56	65	77	47	0.24
$\boldsymbol{f}_{0c}$ augmented with BrI and fever	70	59	67	78	49	0.28
$\boldsymbol{f}_{0c}$ augmented with BrI, fever and Age	73	62	70	80	53	0.34
(B) Results with selected cough-features ($\boldsymbol{f}_{0s}$)						
Selected Cough Features ($\boldsymbol{f}_{0s}$) only	69	64	68	80	50	0.31
$\boldsymbol{f}_{0s}$ augmented with BrI	72	69	71	83	54	0.39
$\boldsymbol{f}_{0s}$ augmented with (existence of) fever	72	60	68	79	51	0.31
$\boldsymbol{f}_{0s}$ augmented with Age	71	67	70	82	53	0.36
$\boldsymbol{f}_{0s}$ augmented with BrI and fever	77	74	76	86	60	0.49
$\boldsymbol{f}_{0c}$ augmented with BrI, fever and Age	80	73	78	86	63	0.51

Results were generated with a leave-one-out validation on the dataset D_{md}. Note that Sens, Spec, Acc, PPV, NPV and κ respectively denote sensitivity, specificity, accuracy, positive predictive value, negative predictive value and the Cohen's kappa statistic.

TABLE 5. The performance of the selected (best) LRM models in diagnosing pneumonia at the patient level, based on the pneumonia cough index (PCI).

	Dataset D_{md}						
Selected LRM model	$\mathrm{PCI}_{\mathrm{ths}}$	Sens (%)	Spec (%)	Acc (%)	PPV (%)	NPV (%)	κ
$\{\mathbf{M}_{\mathrm{cs}}, \Re_{0s}\}$: only cough-features	0.5	93	90.5	92	95	86	0.83
$\{\mathbf{M}_{\mathrm{as}}^{(1)}, \Re_{1s}\}$: cough and BrI	0.5	93	90.5	92	95	86	0.83
$\{\mathbf{M}_{\mathrm{as}}^{(2)}, \Re_{2s}\}$: cough and fever	0.5	91	86	89	93	82	0.76
$\{\mathbf{M}_{\mathrm{as}}^{(3)}, \Re_{3s}\}$: cough and Age	0.5	91	91	91	95	83	0.80
$\{\mathbf{M}_{\mathrm{as}}^{(4)}, \Re_{4s}\}$: cough, BrI and fever	0.5	93	90.5	92	95.5	86	0.83
$\{\mathbf{M}_{\mathrm{as}}^{(5)}, \Re_{5s}\}$: cough, BrI, fever and age	0.5	95.6	90.5	94	95.6	90.5	0.86

Results were obtained *via* a leave-one-out validation technique on the dataset D_{md} Note that $\mathrm{PCI}_{\mathrm{ths}}$, Sens, Spec, Acc, PPV, NPV and κ respectively denote PCI threshold, sensitivity, specificity, accuracy, positive predictive value, negative predictive value and the Cohen's kappa statistic.

technique. In this Section, we explore the performance of the proposed method on our Prospective Validation data set D_{pv}, which is completely independent of D_{md}. The prospective dataset consisted of $C_{pv} = 375$ cough events extracted from $\mathsf{N}_{pv} = 25$ patients. From each patient in D_{pv} we extracted the first 15 coughs available in the recording, following the procedure detailed in "The Development of the Cough Sound Database".

In the Prospective Study we used the LRM models $\{\mathrm{M}_{as,}^{(i)}\Re_{is}\}$, $i = 1,2,\ldots5$, and $\{\mathbf{M}_{cs}, \Re_{0s}\}$ as determined in "Performance of the Proposed Method on the Model Development Dataset D_{md}" section. Decision thresholds PCI_{th} and λ were also fixed to the values given in that section. No further model training or refinement occurred in the Prospective Study.

Table 6 shows the performance of the models in classifying individual cough events into Pneumonic and non-pneumonic classes. The model $\{\mathbf{M}_{cs}, \Re_{0s}\}$, which used only cough-based features, resulted in a sensitivity and specificity of 80 and 63% respectively. The model $\{\mathrm{M}_{as,}^{(5)}\Re_{5s}\}$achieved the best classification performance with a sensitivity = 85.5% and a specificity = 88%.

Next we show the performance of our models in diagnosing pneumonia at the patient level, based on the proposed PCI index. To understand how the number of coughs (P) used to compute the PCI will change the performance of the models, we varied P in (4) from 1 to 13 in steps of 2. For any given P, all possible combinations of P coughs ($^{15}\mathrm{C}_{\mathrm{P}}$ in number) were used in the analysis. The model performance was computed averaged over all $^{15}\mathrm{C}_{\mathrm{P}}$ combinations of coughs per patient. Then the sensitivity and specificity of diagnosis, compared to our reference standard, was evaluated over the dataset D_{pv}.

Figure 4 shows the variation of model performance with P = {1, 3, 5, 7, 9, 11, 13}. According to Fig. 4, the performance of all the models generally increases with an increase in P. However, the performance of the algorithm converges fast as the number of coughs is increased, and 5–11 coughs are sufficient to get the desired level of performance.

Table 7 illustrates the performance of our method in diagnosing pneumonia at the patient level, based on $P = 11$ cough events. The LRM model $\{\mathbf{M}_{cs}, \Re_{0s}\}$, which used only cough-based features, achieved a sensitivity and specificity of 94 and 75%. With $P = 5$ the same model gave us a sensitivity and specificity of 88 and 75% respectively. These results on prospective data lend strong support to the hypothesis that cough sounds carry vital information to diagnose childhood pneumonia.

The top three performing models for the diagnosis of pneumonia were $\{\mathbf{M}_{cs},\Re_{0s}\}$, $\{\mathrm{M}_{as,}^{(2)}\Re_{2s}\}$ and $\{\mathrm{M}_{as,}^{(5)}\Re_{5s}\}$, all with kappa agreement $\kappa > 0.7$, indicating substantial agreement with our reference standard of diagnosis. The model $\{\mathrm{M}_{as,}^{(2)}\Re_{2s}\}$, which used cough-based features augmented with the existence of fever, achieved the highest performance at a sensitivity of 94% and a specificity of 100%.

DISCUSSION AND CONCLUSION

In this paper we proposed an automated algorithm to diagnose childhood pneumonia using cough sounds acquired with microphones that do not require physical contact with patients. The method is based on initially classifying individual cough events into 'pneumonic' and 'non-pneumonic' classes and then calculating a Pneumoic Cough Index (PCI) over all the cough events analysed. Working on 815 cough events from 91 pediatric patients diagnosed with a range of respiratory diseases, we showed that our methods are capable of classifying pneumonia at a sensitivity >90% while holding the specificity at >85%.

Our results are unprecedentedly high and indicate the feasibility of taking a cough-centred approach in diagnosing pneumonia in resourse-poor regions. Furthermore, using features derived from cough sounds only we obtained a sensitivity >90% at a specificity of 75%. The technology, in the simplest version, will require between 5 and 10 cough sounds and will

TABLE 6. Performance of the selected (best) LRM models in classifying individual cough events into classes pneumonic and non-pneumonic. Results were obtained with the Prospective Data Set D_{pv}. Note that Sens, Spec, Acc, PPV, NPV and κ respectively denote sensitivity, specificity, accuracy, positive predictive value, negative predictive value and the Cohen's kappa statistic.

	Prospective dataset D_{pv}					
Selected LRM model	Sens (%)	Spec (%)	Acc (%)	PPV (%)	NPV (%)	κ
$\{\mathbf{M}_{cs}, \Re_{0s}\}$: only cough-features	80	63	75	82	60	0.42
$\{\mathbf{M}_{as}^{(1)}, \Re_{1s}\}$: cough and Brl	80	60	74	81	59	0.4
$\{\mathbf{M}_{as}^{(2)}, \Re_{2s}\}$: cough and fever	83	82.5	83	91	69	0.62
$\{\mathbf{M}_{as}^{(3)}, \Re_{3s}\}$: cough and Age	85	74	82	87.5	70	0.58
$\{\mathbf{M}_{as}^{(4)}, \Re_{4s}\}$: cough, Brl and fever	81	80	81	89.6	67	0.58
$\{\mathbf{M}_{as}^{(5)}, \Re_{5s}\}$: cough, Brl, fever and age	85.5	88	86.4	94	74	0.70

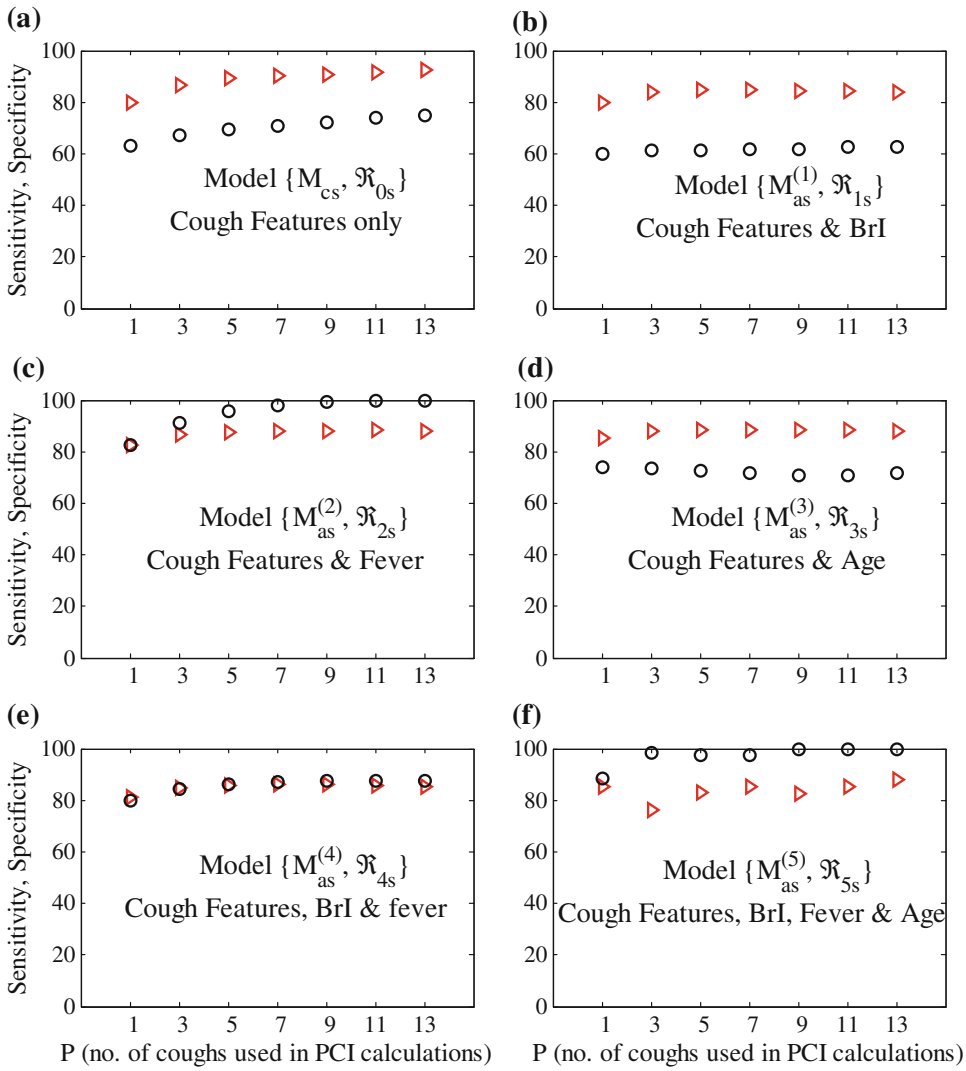

FIGURE 4. The diagnostic sensitivity (triangular markers) and specificity (circular markers) varies with the number of cough events available for analysis. These results were obtained from the Prospective Dataset D_{pv} with LRM models using features as indicated. Five to eleven cough events are generally sufficient to diagnose pneumonia with a high reliability.

TABLE 7. The performance of the selected (best) LRM models in diagnosing pneumonia at the patient level, based on the pneumonia cough index (PCI).

	Prospective dataset D_{pv}					
Selected LRM model	Sens (%)	Spec (%)	Acc (%)	PPV (%)	NPV (%)	κ
$\{\mathbf{M}_{cs}, \Re_{0s}\}$: only cough-features	94	75	88	89	86	0.72
$\{\mathbf{M}_{as}^{(1)}, \Re_{1s}\}$: cough and BrI	88	62.5	80	83	71	0.52
$\{\mathbf{M}_{as}^{(2)}, \Re_{2s}\}$: cough and fever	94	100	96	100	89	0.91
$\{\mathbf{M}_{as}^{(3)}, \Re_{3s}\}$: cough and Age	88	75	84	88	75	0.63
$\{\mathbf{M}_{as}^{(4)}, \Re_{4s}\}$: cough, BrI and fever	82	87.5	84	93	70	0.66
$\{\mathbf{M}_{as}^{(5)}, \Re_{5s}\}$: cough, BrI, fever and age	88	100	92	100	80	0.83

Results were obtained with the Prospective Data Set D_{pv}. $P = 11$ cough events were used for PCI calculation in Eq. (4). Note that Sens, Spec, Acc, PPV, NPV and κ respectively denote sensitivity, specificity, accuracy, positive predictive value, negative predictive value and the Cohen's kappa statistic.

automatically and immediately provide a diagnosis, without requiring physical sensor contact with patients. Such a system, if successful, is expected to be a paradigm shifting novelty in the field of pneumonia diagnosis in remote regions. It may also find applications in such areas as respiratory monitoring in elderly care facilities throughout the world.

The simplicity of the proposed technology and potential for low cost implementations on ubiquitous devices make our approach valuable. It may have strategic value in developing new vaccines as well as management strategies for childhood pneumonia. Clinical trials of new pneumonia vaccines conducted in resource-limited regions of the world require reliable tools to measure the efficacy of intervention. The same is true for assessing the effectiveness of any new pneumonia management strategy targeting large populations. There are no field deployable gold standards to diagnose pneumonia, especially the early stage (non-severe) disease where even chest X-rays fail.[3] The existing WHO algorithm is limited due to its low diagnostic performance. While the algorithm serves a very useful role in picking up potential cases of pneumonia, the cost of doing so, the low specificity, makes it difficult to use in assessing the validity of a new intervention.

During a cough event, lungs are connected to the atmosphere *via* a column of air, which, we hypothesise, can support a much higher bandwidth than the traditional pathway across the chest musculature. In this paper, we propose to use this ''Information Super Highway'' to diagnose pneumonia.

Physics dictate that sounds generated inside the lungs, including disease-specific ones, should propagate outside through the air column at the speed of sound. In pneumonia, lung consolidation and fluid in lower airways too modulate the sounds measured outside. However, it is likely that these components get buried in the loud main component of a cough contributed by the vibration of upper airway.

Current knowledge on cough is inherently limited by the boundaries of human perception: (a) the average middle-aged adult cannot hear above 15 kHz, (b) humans cannot hear below the amplitude threshold of hearing, (c) louder sounds can substantially increase the threshold of hearing of fainter sounds, which is known as the ''*masking effect*''. Subjective assessment also limits the amount of information extractable from cough in a clinical diagnosis. Our methods attempt to surpass these limitations.

As far as we know, this is the first attempt in the world to develop an objective model for the penumonia diagnosis in humans centred about cough sounds. Our technology may also have applications in diagnosing respiratory diseases even in the developed world, in a primary care clinical setting where immediate access to laboratory facilities is not available.

While the ability to use cough sounds in diagnosing pneumonia is a groundbreaking outcome, our approach also allows the integration of other features to further increase the diagnostic performance of the method, when necessary. For instance, the inclusion of the *existence of fever* (as a binary variable assuming values: 'yes' or 'no'), increases the specificity of the cough-based algorithm by 33% while keeping the sensitivity at the high value of 94%. Currently we are working on implementing our technology on smart phones and also in building dedicated low-cost electronics devices incorporating several other clinical measurements.

The results presented in this paper may be improved by systematically controlling the recording environment when possible. The recording environment and background noise may affect the performance of the algorithm. We are greatly encouraged, however, that the real-world measurements we did in a respiratory ward in Indonesia so far have led to unprecedented outcomes exceeding those of existing WHO algorithms. In the future, the robustness of the algorithm should be further explored.

Our methods have not yet been clinically tested in a trial involving a large number of subjects from a community-based population. In addition, the reference standard used in diagnosing pneumonia is the overall clinical diagnosis aided by laboratory testing, X-ray asessment (when clinically required) and the clinical course of the disease. Not all subjects have gone though X-ray imaging. Results presented in this paper should be interpreted in this context.

APPENDIX

Our method requires the computation of a number of mathematical features from cough sounds. This Appendix describes the features we computed from each sub-segment x_i, $i = 1, 2, 3$ of a recorded cough sound x.

The Bispectrum Score (BGS)

The 3rd order spectrum of a signal is known as the bispectrum.[2] Unlike the power spectrum (the 2nd order spectrum based on the autocorrelation), the bispectrum preserves Fourier phase information. The bispectrum $B_{xi}(\omega_1, \omega_2)$ of the segment x_i can be estimated from (5) as,

$$B_{xi}(\omega_1, \omega_2) = \sum_{\tau_1=-\infty}^{\tau_1=+\infty} \sum_{\tau_2=-\infty}^{\tau_2=+\infty} W(\tau_1, \tau_2) \cdot C_{xi}(\tau_1, \tau_2) e^{-j(\tau_1\omega_1+\tau_2\omega_2)} \quad (5)$$

where $W(\tau_1, \tau_2)$ is a bispectrum window function such as the minimum bispectrum-bias supremum window[14] used in this paper, $C_{xi}(\tau_1, \tau_2)$ is the third order cumulants of x_i estimated with (6), and, ω_1, ω_2 denotes digital frequencies.

$$C_{xi}(\tau_1, \tau_2) = \frac{1}{L}\sum_{k=0}^{L-1} x_i(t)x_i(t+\tau_1)x_i(t+\tau_2), | \quad \tau_1| \le Q, |\tau_2| \le Q \tag{6}$$

In (6), Q is the length of the 3rd order correlation lags considered and x_i is a zero-mean signal.

The bispectrum is a 2D signal. However, it can be proven[2] that for linear signals, any 1D oblique slice of the bispectrum other than the slices parallel to the axes: $\omega_1 = 0$, $\omega_2 = 0$ and $\omega_1 + \omega_2 = 0$ carries sufficient information to characterise the entire 2D bispectrum within a phase factor. In this work, we capture the information available in the bispectrum *via* the diagonal slice P(ω) defined by $\omega_1 = \omega_2 = \omega$, i.e. $P(\omega) = B_{xi}(\omega, \omega)$.Then the Bispectrum Score (BSG) is computed as defined in (7). In (7) we used $k_1 = 90$ Hz, $k_2 = 5$ kHz, $k_3 = 6$ kHz and $k_4 = 10.5$ kHz.

$$BSG = \frac{\int_{k1}^{k2} |P(\omega)| \cdot d\omega}{\int_{k3}^{k4} |P(\omega)| \cdot d\omega} \tag{7}$$

Non-Gaussianity Score (NGS)

NGS score is a numerical measure of non-Gaussianity of a given segment of data x_i. The normal probability plot can be utilized to obtain a visual measure of the Gaussianity of a set of data, and the NGS score is a way of quantifying the non-Gaussianity based on regression analysis. We used (8) to estimate the NGS score, where p and q represents the normal probability plots of the reference normal data and the analysed data (x_i). The symbol N is the number of data points used in the probability plot.

$$NGS = 1 - \left(\frac{\sum_{j=1}^{N}(q[j] - p)^2}{\sum_{j=1}^{N}(q[j] - \bar{q})^2}\right) \tag{8}$$

Formant Frequencies

In speech analysis, formants frequencies (FF) are referred to as the resonances of the vocal tract.[16] In cough analysis, it is reasonable to expect that the resonances of the overall airway that contribute to the generation of a cough sound will be represented in the formant structure. One classic example for this is wheeze. Existence of mucous can also change acoustic properties of airways. We included the first four formants (F1, F2, F3, F4) in our candidate feature set. We computed F1–F4 by peak picking the Linear Predictive Coding (LPC) spectrum of cough segments x_i. For this work we used a 14th order LPC model with the parameters determined *via* the Levinson-Durbin recursive procedure.[17]

Log Energy (LogE)

The log energy for every sub-segment x_i was computed using (9):

$$\text{LogE} = 10\log_{10}\left(\varepsilon + \frac{1}{N}\sum_{k=1}^{K}\left(x_i(t)^2\right)\right) \tag{9}$$

In (9) ε is an arbitrarily small positive constant added to prevent any inadvertent computation of the logarithm of 0.

Zero Crossing (Zcr)

The number of zero crossings was counted for each sub-segment x_i.

Kurtosis (Kurt)

The kurtosis is a measure of how peaky the probability density distribution of x_i is. It is the fourth central moment of x_i and can be computed using (10), where μ and σ respectively denote the mean and the standard deviation of x_i.

$$\text{Kurt} = \frac{E\{(x_i[k] - \mu)^4\}}{\sigma^4} \tag{10}$$

Mel-Frequency Cepstral Coefficients (MFCC)

MFCCs have been widely used in speech recognition systems.[7,9] MFCC provides some resilience to the non-linguistic sources of variance in speech signals. They also provide orthogonal features making facilitating the training of the classifier. The computation of MFCC involves the estimation of short-term power spectra, mapping to Mel frequency scale and then computing the cepstral coefficients. In our work, we included 12 MFCC coefficients in our feature set.

ACKNOWLEDGMENTS

This work was supported by the Bill & Melinda Gates Foundation, USA, under its Grand Challenges in Global Health Explorations Grant program (Grant ID# OPP1008199 GCE to Abeyratne).

REFERENCES

[1]Abaza, A. A., J. B. Day, J. S. Reynolds, A. M. Mahmoud, W. T. Goldsmith, W. G. McKinney, E. L. Petsonk, and D. G. Frazer. Classification of voluntary cough sound and airflow patterns for detecting abnormal pulmonary function. *Cough* 5:8, 2009.

[2]Abeyratne, U. Blind Reconstruction of Non-minimum-phase Systems from 1-D Oblique Slices of Bispectrum. IET, 1999. .

[3]Bhutta, Z. A. Childhood pneumonia in developing countries: refinement of clinical algorithms is a priority. *BMJ* 333:612, 2006.

[4]Cardoso, M. R. A., C. M. Nascimento-Carvalho, F. Ferrero, F. M. Alves, and S. N. Cousens. Adding fever to WHO criteria for diagnosing pneumonia enhances the ability to identify pneumonia cases among wheezing children. *Arch. Dis. Child.* 96:58, 2011.

[5]Falade, A., H. Tschäppeler, B. M. Greenwood, and E. K. Mulholland. Use of simple clinical signs to predict pneumonia in young Gambian children: the influence of malnutrition. *Bull. World Health Organ.* 73:299, 1995.

[6]Frosh, A. Pediatric airway: cry, stridor, and cough. *JAMA* 301:2392–2392 (2009).

[7]Gold, B., N. Morgan, and D. Ellis. Speech and Audio Signal Processing: Processing and Perception of Speech and Music. Wiley-Interscience, 2011.

[8]Harari, M., V. Spooner, S. Meisner, M. Carney, F. Shann, and J. De Campo. Clinical signs of pneumonia in children. *Lancet* 338:928–930, 1991.

[9]Hirschberg, J., I. Lelli, and Z. Garay. The Pediatric Airway: Cry, Stridor, and Cough. Plural Pub., 2009.

[10]IMCI Integrated Management of Childhood Illness. World Health Organization, 2005.

[11]Knocikova, J., J. Korpas, M. Vrabec, and M. Javorka. Wavelet analysis of voluntary cough sounds in patients with respiratory diseases. *J. Physiol. Pharmacol.* 59:331–340, 2008.

[12]Kosasih, K., U. Abeyratne, and V. Swarnkar. High frequency analysis of cough sounds in pediatric patients with respiratory diseases. In: Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE. 2012. IEEE.

[13]Lozano, J., M. Steinhoff, J. Ruiz, M. Mesa, N. Martinez, and B. Dussan. Clinical predictors of acute radiological pneumonia and hypoxaemia at high altitude. *Arch. Dis. Child.* 71:323–327, 1994.

[14]Mendel, J. M. Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications. *Proc. IEEE* 79:278–305, 1991.

[15]Mulholland, E., E. Simoes, M. Costales, E. McGrath, E. Manalac, and S. Gove. Standardized diagnosis of pneumonia in developing countries. *Pediatr. Infect. Dis. J.* 11:77, 1992.

[16]Ng, A. K., T. S. Koh, E. Baey, T. H. Lee, U. R. Abeyratne, and K. Puvanendran. Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea? *Sleep Med.* 9:894–898, 2008.

[17]Oppenheim, A. V., R. W. Schafer, and J. R. Buck. Discrete-Time Signal Processing. Vol. 1999. Englewood Cliffs, NJ: Prentice Hall, 1989.

[18]Palafox, M., H. Guiscafré, H. Reyes, O. Muñoz, and H. Martínez. Diagnostic value of tachypnoea in pneumonia defined radiologically. *Arch. Dis. Child.* 82:41–45, 2000.

[19]Piirilä, P., and A. Sovijärvi. Differences in acoustic and dynamic characteristics of spontaneous cough in pulmonary diseases. *Chest* 96:46–53, 1989.

[20]Pneumonia Strategy Overview. Bill & Melinda Gates Foundation (November 2009).

[21]Smith, J. A., J. E. Earis, and A. A. Woodcock. Establishing a gold standard for manual cough counting: video versus digital audio recordings. *Cough* 2:6, 2006.

[22]Suttorp, N., T. Welte, and R. Marre. Community-Acquired Pneumonia. Springer, 2007.

[23]Swarnkar, V., U. Abeyratne, A. Chang, Y. Amrulloh, A. Setyati, and R. Triasih. Automatic identification of wet and dry cough in pediatric patients with respiratory diseases. *Ann. Biomed. Eng.* 41, 2013. doi:10.1007/s10439-013-0741-6.

[24]Wardlaw, T., E. W. Johansson, and M. Hodge. Pneumonia: The Forgotten Killer of Children. UNICEF/WHO, 2006.

[25]Thorpe, W., M. Kurver, G. King, and C. Salome. Acoustic analysis of cough. In: Intelligent Information Systems Conference, The Seventh Australian and New Zealand 2001. 2001. IEEE.

[26]WHO/UNICEF. Global Action Plan for the Prevention and Control of Pneumonia (GAPP). WHO Reference Number: WHO/FCH/CAH/NCH/09.04, 23 (2009).