

RLC-S

Recognition and Localization of a Crackle Sound

Providing a clinical tool for assessing in crackle sound detection in a respiratory physiotherapy for Cystic Fibrosis Patients

Akiva Hassan	iqab.hassan@mail.huji.ac.il
Prof. Yuval Kochman	yuval.kochman@mail.huji.ac.il
Mr. Adi Szekin	adi.szes@gmail.com
Jana awdy	jana.awdy@mail.huji.ac.il

Abstract

Respiratory system diseases affect people's social, economic and health life significantly. For these reasons, a lot of researches are going on for early diagnosis and intervention in respiratory diseases. Chronic respiratory diseases (CRDs) an inherited disorder that causes severe damage to the lungs, digestive system, and other organs in the body, finally the disorder limits the ability to breathe over time. This disorder is the second most common life-shortening, childhood-onset genetic disease.

Lung sound characteristics provide important clues in the diagnosis of respiratory abnormalities and infection. While using the stethoscope is the most common way to diagnose the abnormalities it is not accurate enough, takes a lot of time and is dependable on the experience of the physician.

The goal of this project is to provide a semi-automated detection health system tool of crackle and wheezes sounds using an acoustical sensor. The sensor will optimize and shorten treatment time, using state of the art, signal processing and machine learning algorithms to recognition lung sound. Our project is semi- automated tool that will help physician and physiotherapist.

Introduction

Chronic respiratory diseases (CRDs) are among the principal causes of mortality and morbidity around the world, according to the World Health Organization [1]. The first approach employed in the diagnosis of pulmonary diseases is the clinical examination of the pulmonary function that includes clinical history and auscultation of the lungs with the stethoscope. During the auscultation procedure, adventitious lung sounds added on the breath or base lung sounds are a common finding. Although the stethoscope remains the most widely used instrument in clinical medicine and still guides diagnosis when other pulmonary function tests are not available, the auscultation by the stethoscope has several limitations, e.g., it is a subjective process that depends on the ability and expertise of the physician [2], it is limited by human audition [3], the stethoscope may be more adequate for cardiac auscultation [4], and the lung sounds are not permanently recorded for further analysis.

In this context, lung sound characteristics provide important clues in the diagnosis of respiratory abnormalities and infections. Auscultation is an effective technique in which physicians evaluate and diagnose the disease after using a stethoscope for lung disease. Tis method is both inexpensive and easy, and it does not require internal intervention into the human body [5]. However, traditional stethoscopes may be exposed to external noise sounds, weaken the sound components above 120 Hz, and cannot filter the audio frequencies of the body in auscultation and cannot create permanent recordings in monitoring of the disease course [5]. In addition, accurate diagnosis of diseases requires highly experienced medical staff. Therefore, it is

important to use electronic instrumentation and systems which operate with artificial intelligence and pattern recognition to assist doctors in decision making process. As a result, it is practically contributed to a specialist who works under stress, fatigue, and intensive conditions

Nowadays, it is recognized that respiratory sounds (RS) make it possible to obtain information about the respiratory health of a subject in a noninvasive fashion, e.g., the characteristics of RS differ between different pulmonary disorders, reflecting different pathophysiology's and severity levels [6].

Related works

Many diagnosing techniques were suggested beside factors that can distinguish a crackle and wheezes sound from another. Starting from the frequencies, normal breath sounds are noises with frequencies generally between 200 and 600 Hz while crackles and wheezes are a sequence of explosive, nonmusical, interrupted pulmonary sounds with a wide spectrum of frequencies between 400 and 2500 Hz [7].

Crackles and wheezes also characterized by their specific waveform, their duration, and their location in the respiratory cycle [7].

In [8], two datasets namely continuous adventitious sound (CAS) and tracheal breath sound (TBS) were considered. TBS and CAS datasets were further divided into two sections: inspiratory and expiratory. TBS and CAS dataset have the following class labels; wheezing, stridor, rhonchi and mixture lung sounds. Distinction function, instantaneous kurtosis, and SampEn were used for feature extraction. The reported accuracy scores were in the range of 97.7% and 98.8% that were obtained with SVM classifier using the Radial Basis Function (RBF) kernel. In [9], the authors used the ICBHI 2017 challenge database which has normal, wheezes, crackles and wheezes plus crackles class labels. the ICBHI 2017 is a challenging database, since there are noises, background sounds and different sampling frequencies (4 kHz, 10 kHz, 44.1 kHz). In [7], spectral features and Decision Tree were chosen for feature extraction and classification, respectively. In [10], authors chose short time Fourier transform (STFT) and STFT+Wavelet to extract features and principal component analysis (PCA) to reduce the process load while testing the algorithm performance with the SVM classifier. In this paper, it was worked to boost the classification performance for ICBHI 2017 database which is quite challenging. In this context, spectrogram images were utilized to create time–frequency transformation from the lung sounds. These spectrogram images were used as input to the deep feature extraction and transfer learning. SVM and softmax classifiers were used for deep features and transfer learning approaches, respectively. the performances of proposed methods are evaluated by accuracy, sensitivity and specificity scores. the results were also

compared with some of the existing results. the proposed schemes improved the classification performance of the lung sound discrimination.

Methods & Materials

The proposed methods for lung sound classification are shown in Figs. 1 and 2, respectively. The pre-trained VGG-16 model is considered for both deep feature extraction and fine-tuning. As seen in Figs. 1 and 2, the proposed methods initially convert the input lung sound signals into time–frequency images. The short time Fourier transform (STFT) is used for T-F image construction. Because lung sounds are recorded at different frequencies, the window sizes that should be used for the STFT are different. Window sizes are chosen between 0.01 and 0.025 times of the sampling frequency because they would better reveal the lung sound characteristics [12]. After the T-F images are constructed, they are resized to 224×224 (VGG16) because of being suitable with deep feature extraction and transfer learning. The fully connected layers whose outputs are 4096-dimensional are used for deep feature extraction. The pre-trained VGG16 model was chosen for transfer learning in Fig. 2. The pre-trained VGG16 model. is further trained with the input lung spectrogram images called fine-tuning. Since layers are configured for 1000 classes of ImageNet challenge and consider that normally, ImageNet is recognizing objects. We have spectrogram images which has signals. So we set 20 out of 23 layers to trainable.

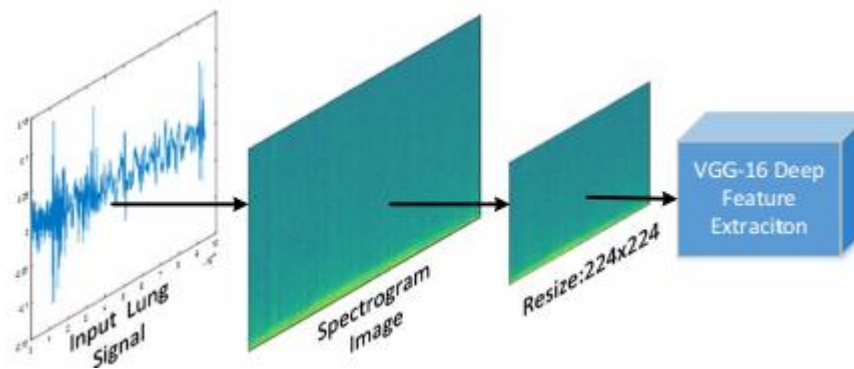


Figure 1 The proposed deep feature extraction methodology for lung sound classification

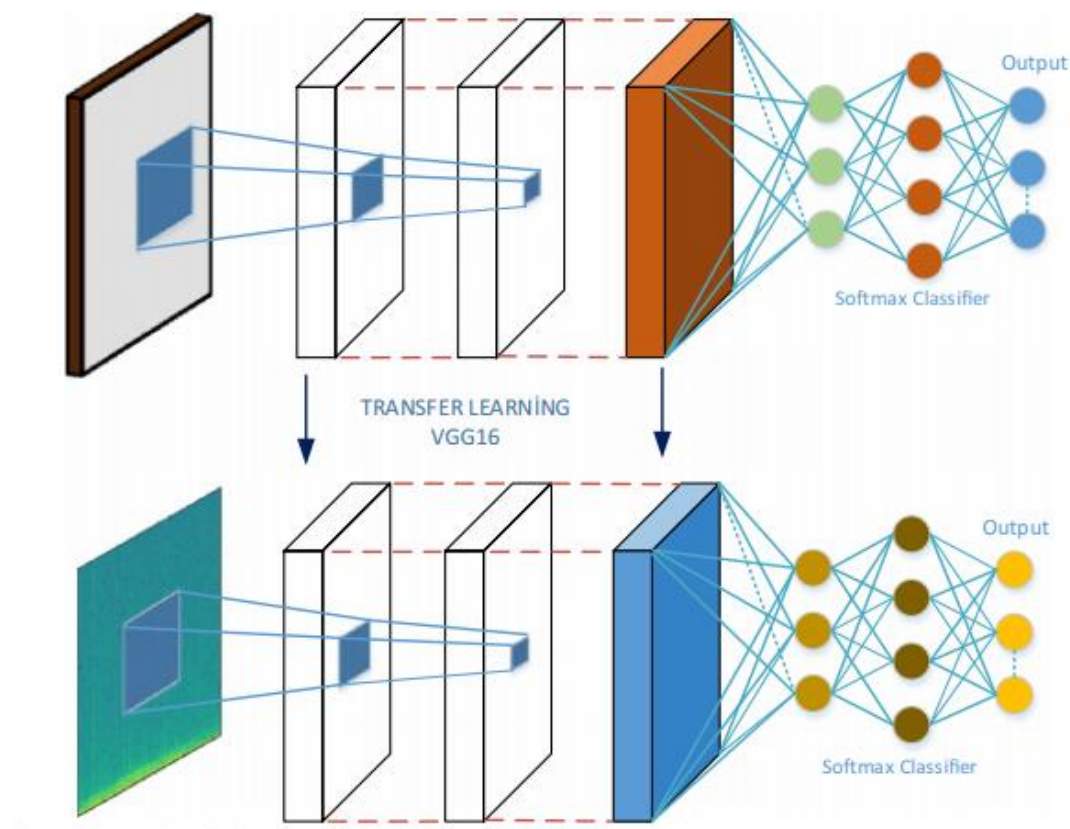


Figure 2 The proposed transfer learning methodology for lung sound classification

Description of Database

The Respiratory Sound Database was created by two research teams in Portugal and Greece. It includes 920 annotated recordings of varying length - 10s to 90s. These recordings were taken from 126 patients. There is a total of 5.5 hours of recordings containing 6898 respiratory cycles - 1864 contain crackles, 886 contain wheezes and 506 contain both crackles and wheezes. The data includes both clean respiratory sounds as well as noisy recordings that simulate real life conditions. The patients span all age groups - children, adults, and the elderly.

The labelling details of a sample lung sound is given in Table 1. the 20-s audio file in Table 1 is divided into 9 cycles according to the start and end times. The number of these cycles and the sampling frequency (4 kHz, 10 kHz, 44.1 kHz) are different for each audio file. The class labels of the divided cycles are set by checking the values of wheezes and crackles columns in Table 1. The class labels are crackles, wheezes, normal and wheezes plus crackles if the wheezes and crackles values are 1-0, 0-1, 0-0

and 1-1. Also, the total number of ICBHI 2017 dataset cycles is shown to class labels in Table 2.

Table 1: Cycle info for an audio file of ICBHI 2017 database

Cycles	Start time	End time	Crackles	Wheezes
1	0.634	2.985	0	0
2	2.985	4.998	0	1
3	4.998	7.684	0	1
4	7.684	10.223	0	1
5	10.223	13.002	1	1
6	13.002	14.945	1	1
7	14.945	17.476	1	0

Table 2: The total number of ICBHI 2017 dataset cycles

Dataset	Number total
Number of normal cycles	3642
Number of cycles with crackles	1864
Number of cycles with wheezes	886
Number of cycles with both	506
Number total of cycles	6898

Spectrogram:

The Short Time Fourier Transform (STFT) represents the frequency content of a (windowed) audio segment. Concatenating such representations over time allows for a waveform to be visualized as a 2-D function of time and frequency.

For spectrogram creation, we used Hamming window with 1024 ms and the number of the FFT was chosen as 3000. The window-overlap sizes were selected as 512-64 for 44.1 kHz sampling frequency, 128-16 for 10 kHz sampling frequency and 64-8 for 4 kHz sampling frequency Fig[3].

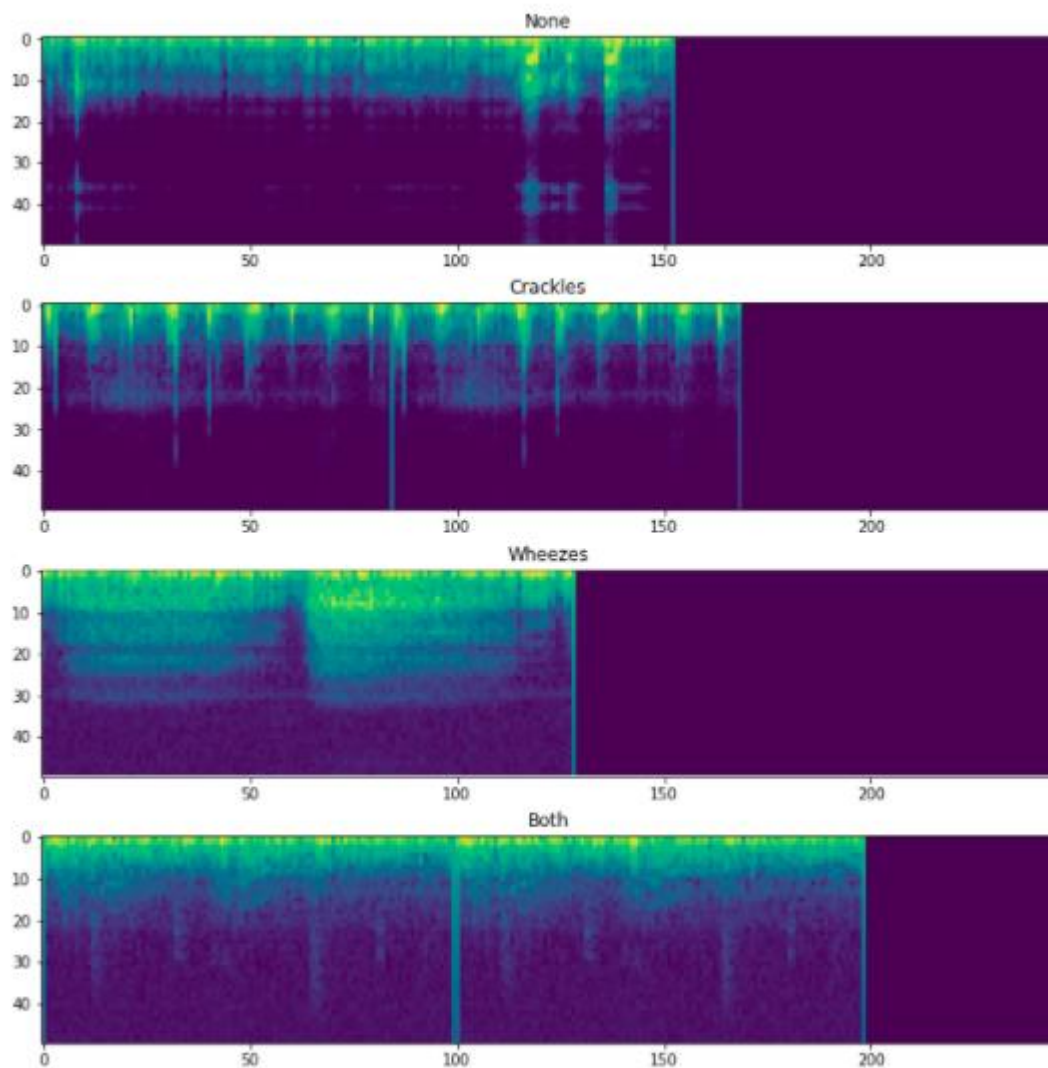


Figure 3 Short Time Fourier Transform plot for different abnormal birthing noise.

Convolutional neural networks

(CNN) Convolutional neural networks (CNNs) take place as the category of deep neural networks, in terms of searching application to present categorization of images and analysis [10]. The categorization and attribute extraction are given by end-to-end learning architecture of CNNs. CNNs are composed of convolution, pooling, and fully connected layers. the pooling and convolution layers are back-to-back utilized for the formation of network architecture and building high degree of discernable feature set for categorization. Categorization performing which use feature set obtained from previous layers is demanded for the fully connected layers. The traditional back propagation algorithm updates a great number of parameters in CNN model training.

A non-linear activation function such as ReLU ($R(z) = \max(0, z)$) is used to the feature map that is constituted by convolution operation. The aim of the maxpooling layer is to conjugate semantically convenient features came from the previous layer. the max-pooling layer executes down-sampling operation by splitting the previous layer into rectangular pooling regions and calculating the maximum value of each region [19]. Fully connected layer, softmax classifier are existed in last stage of CNNs. The fully connected layer has a transmission mission between previous layer and classification layer. The fully connected layers can be summarized in three stages. In the first stage, the results of convolution and pooling layer are fattened and converted them to a column vector that will be an input layer for the latter stage. In the second stage, inputs from previous stage are taken for the feature analysis, and the weights are applied to predict the true labels. In the last stage, the final prediction scores for each class label is determined.

Transfer learning

In transfer learning (TL), which is contemporary trend in deep learning, the layers of a pre-trained network are shared or conveyed to other networks for fine-tuning or features extraction [12]. Initially, the training of a CNN model is carried out by a large dataset. After this process, training of pre-trained model is conducted once more by a smaller dataset to get fine tuning for developing estimated performance of CNN model. TL which provides satisfying tuning is more enduring than CNN model training from scratch. While initial layers represent features such as curves, color blobs, edges in CNN architecture, abstract and specific features are provided by final layers [16, 17].

SoftMax classifier

The generalized binary form of logistics regression is used for the SoftMax classifier. Like hinge loss functions, g which is mapping function is the linear dot product of x_i symbolized as input data and of ω symbolized as weight matrix, as shown in Eq. 1 [13].

$$g(x_i, \omega) = \omega \cdot x_i \quad (1)$$

VGG16 Architecture:

The input to conv1 layer is of fixed size 224×224 RGB image. The image is passed through a stack of convolutional (conv.) layers, where the filters were used with a very small receptive field: 3×3 . In one of the configurations, it also utilizes 1×1 convolution filters, which can be seen as a linear transformation of the input channels. The convolution stride is fixed to 1 pixel; the spatial padding of conv. layer input is such that the spatial resolution is preserved after convolution, i.e. the padding is 1-pixel for 3×3 conv. layers. Spatial pooling is carried out by five max-pooling layers, which follow some of the conv. layers (not all the conv. layers are followed by max-pooling). Max-pooling is performed over a 2×2 -pixel window, with stride 2.

Three Fully-Connected (FC) layers follow a stack of convolutional layers (which has a different depth in different architectures): the first two have 4096 channels each, the third performs 1000-way ILSVRC classification and thus contains 1000 channels (one for each class). The final layer is the soft-max layer. The configuration of the fully connected layers is the same in all networks.

Since layers are configured for 1000 classes of ImageNet challenge, the last three layers of the VGG16 model are not taken into account in order to get fine-tuning adaptation.

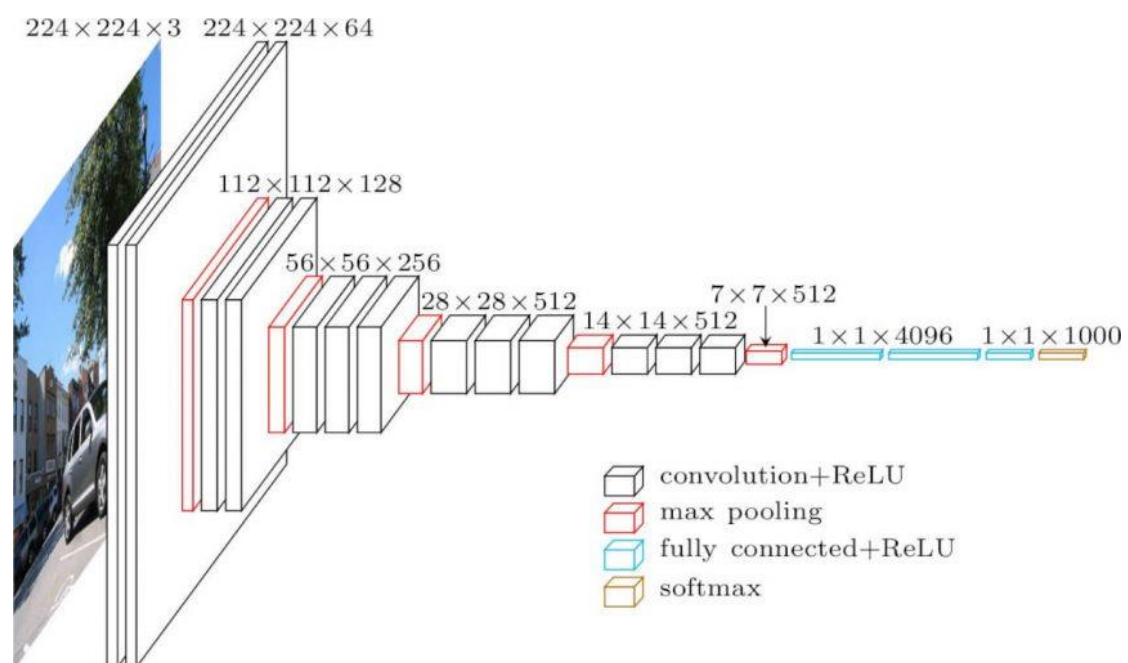


Figure 4 VGG16 Architecture

Evaluation and verification:

The Accuracy (ACC), Sensitivity (SN), Specificity (SP) and Precision (PREC) have been used as a performance metrics to evaluate CNN models efficiency. These performance metrics are shown in Eq.[2-5]

$$\bullet \text{ ACC} = \frac{TP + TN}{TP + TN + FN + FP} = \frac{TP + TN}{P + N} \quad (2)$$

$$\bullet \text{ SP} = \frac{TN}{TN + FP} = \frac{TN}{N} \quad (3)$$

$$\bullet \text{ SN} = \frac{TP}{TP + FN} = \frac{TP}{P} \quad (4)$$

$$\bullet \text{ PREC} = \frac{TP}{TP + FP} \quad (5)$$



Figure 5 the result of the the modle acc and Loss

Now we have a set of observed labels and four sets of predicted labels. The set of observed labels is compared with a set of predicted labels to create a confusion matrix. there Four classes need to be tagged: normal\clear breathing, crackles, wheezes, crackles and wheezes. And to best evaluate the accuracy of our model we did two evaluation methods:

- classification for the above four classes
- classification for two classes, sick breathing that include crackles, wheezes, crackles and wheezes class, and normal\clear breathing.

In fig [6] confusion matrix which is a performance measurement for classification was shown for four classes. It has been shown different combinations for all the classes in the confusion matrix, which it is the numbers of true positive, false positive, true negative and false negative.

We can conclude that the model recognize and generalizes wheezes, crackles and wheezes plus crackles and healthy birthing very well. Our result on target sound is acc 0.763.

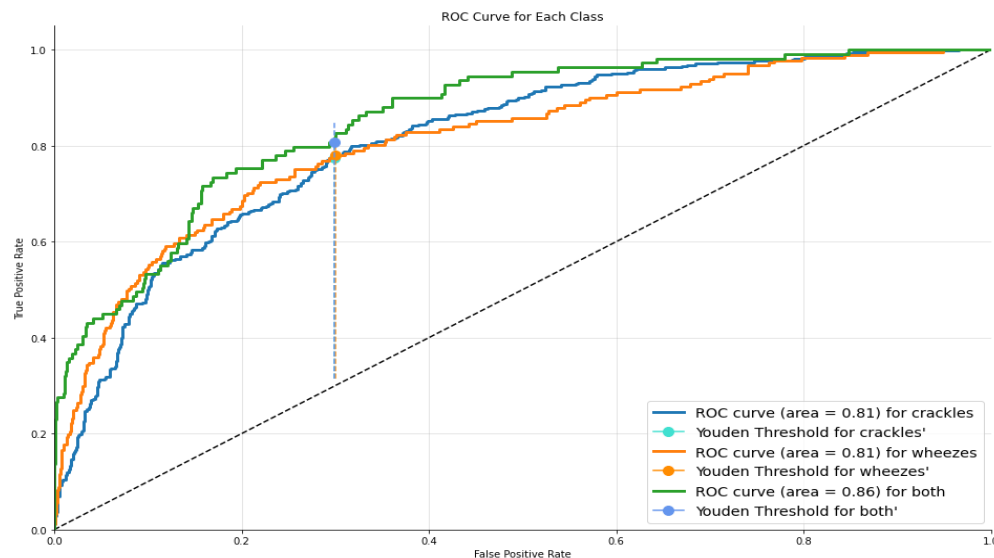


Figure 5 ROC plot to the four classes performance measure of the whole negative part of a dataset, whereas sensitivity is a performance measure of the whole positive part. And the measures the entire two-dimensional area underneath the entire ROC curve. From the ROC plot, we calculate the confusion matrix threshold for the two classes classification. We computed ROC curve and ROC area for each

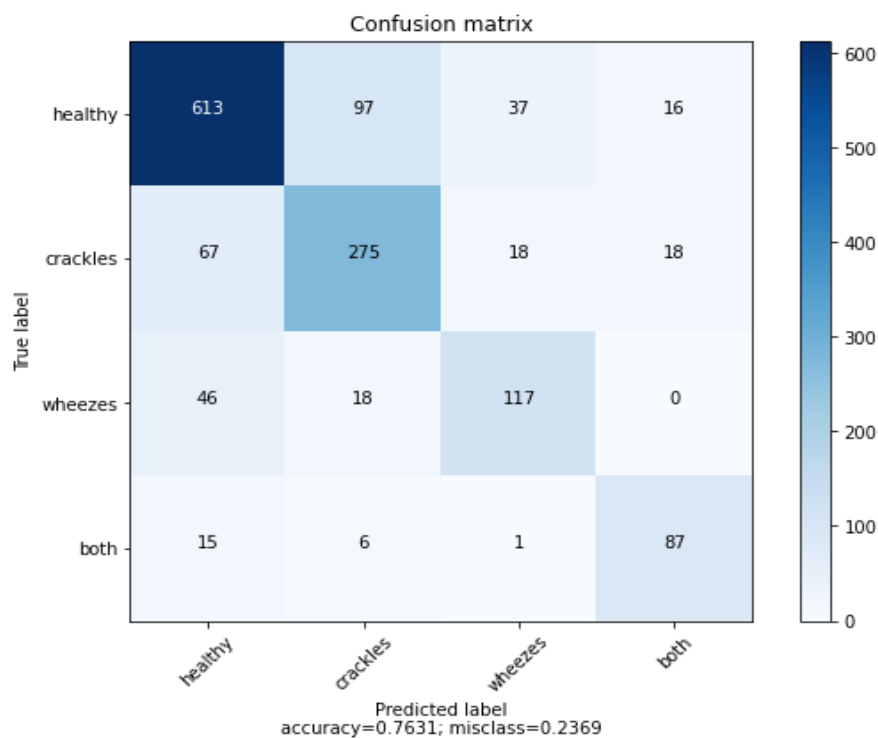


Figure 6 confusion matrix that allowa visualization of the performance of our model, for four classes

We can conclude that the model recognizes and generalizes normal and abnormal birthing, healthy and sick breathing. Since the original data classifies to four classes we calculate the threshold of the confusion matrix to give us the smallest False negative (FN): incorrect negative prediction because our model maybe will used as a medical tolls, to calculate that threshold we use the Roc curve fig[5] ,Our result on target sound is acc 0.793.

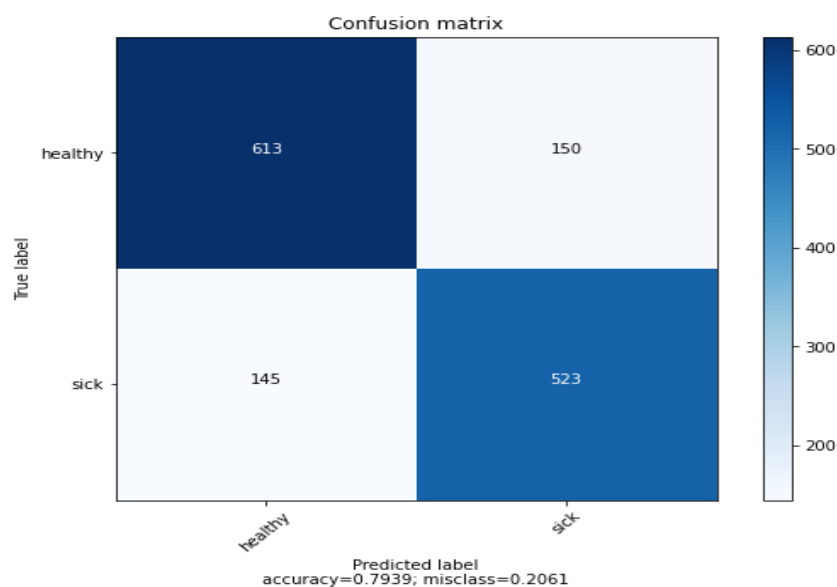


Figure 7 confusion matrix that allowa visualization of the performance of our model, for two classes healthy and sick the the threshold of the matrix was concluded from the

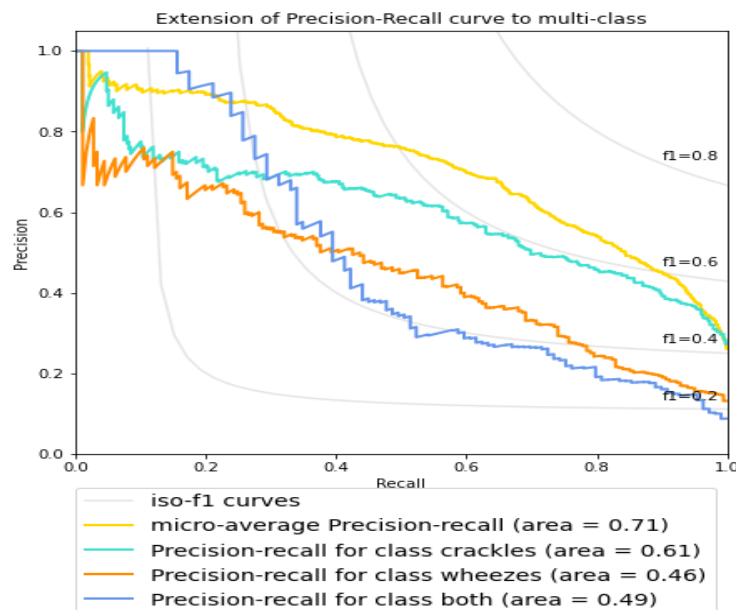


Figure 8 model-wide evaluation measure that is based on two basic evaluation measures – recall and precision. Recall is a performance measure of the whole positive part of a dataset, whereas precision is a performance measure of positive predictions.

Conclusions

In this paper, we focus on the automatic diagnosis of lung diseases which is one of the most important issues in public health for the task of automated detection of adventitious discontinuous respiratory sound like crackles and wheezes. There have been many studies on this subject in the literature, but no challenging data sets including background sounds, noises and different sampling frequencies have been used for lung sound classification. Also most of the work consists of traditional methods. In recognition problem of lung sounds, deep learning, which is state of the art method, is handled to boost the classification performance. In the preprocessing stage of the proposed methods, images that meet one-to-one spectrogram properties were obtained with colormap to extract deep feature and apply fine tuning. In deep learning method, the VGG-16 model of CNN was used to perform feature extraction. the VGG-16 model was preferred for the proposed methods because it gave good classification accuracy. The classification accuracies have been significantly improved for ICBHI 2017 Database containing lung sounds that are difficult to classify.

Finally, we emphasize a model system was intended to provide an accessible tool to quantitatively analyze respiratory sounds while maintaining the desired properties of the auscultation like being non-invasive, ubiquitous, low-cost, and easy-to-use. In

particular, the mobility characteristic of the model allows one to record respiratory sounds at the bedside of the patients without the need to move them, for example, to a specialized laboratory with an anechoic chamber. We consider that this and similar efforts will enable the acquisition of large samples of data with the benefits of a better understanding of pulmonary pathologies at early stages of the disease which in turn can help to deal with the underestimation of respiratory diseases among the general population that without the availability of mobile solutions found difficult to access, or access to late, to the specialized levels of the healthcare system.

Future Work

Current work involves the implementation of signal processing techniques and ML that complement the information obtained from the waveform analysis by the physician. Future work will involve:

- a mobile health system for the automated detection of crackle sounds comprised by an acoustical sensor, a smartphone device, and a mobile application
- to add microphones to count with spatial information. In addition, given the mobility characteristics of the system, it is being used to record a greater number of recordings in the clinical setting, directly at the bedside of the patients, which will be useful to develop more robust detection algorithms as well as to analyze them.
- improvement in the detection of coarse crackles by evaluating other techniques as independent component analysis and high-resolution time-frequency analysis
- the development of a wireless version of acoustical sensor that will enable data acquisition during a diversity of maneuvers involving conditions not restricted to stationary subjects.

Bibliography Referenced:

1	WHO. WHO Chronic Respiratory Diseases (CRDs). 2018. Available online: http://www.who.int/respiratory/en/ (accessed on 24 May 2018)
2	Brooks, D.; Thomas, J. Interrater reliability of auscultation of breath sounds among physical therapists. <i>Phys. Ther.</i> 1995, 75, 1082–1088
3	Piirila, P.; Sovijarvi, A.R. Crackles: Recording, analysis and clinical significance. <i>Eur. Respir. J.</i> 1995, 8, 2139–2148.

4	Pasterkamp, H.; Kraman, S.S.; Wodicka, G.R. Respiratory sounds: Advances beyond the stethoscope. <i>Am. J. Respir. Crit. Care Med.</i> 1997, 156, 974–987
5	Nandini S, Sahidullah M, Goutam S. Lung sound classification using cepstral-based statistical features. <i>Comput Biol Med.</i> 2016;75:118–29
6	Reichert, S.; Gass, R.; Brandt, C.; Andres, E. Analysis of respiratory sounds: State of the art. <i>Clin. Med. Circ. Respir. Pulm. Med.</i> 2008, 2, 45–58
7	Gupta P, Moghimi MJ, Jeong Y, Gupta D, Inan OT, Ayazi F. Precision wearable accelerometer contact microphones for longitudinal monitoring of mechano-acoustic cardiopulmonary signals. <i>NPJ Digit Med.</i> 2020;3:19. Published 2020 Feb 12. doi:10.1038/s41746-020-0225-7
8	. Jin F, Sattar F, Goh DYT. New approaches for spectro-temporal feature extraction with applications to respiratory sound classification. <i>Neurocomputing.</i> 2014;123:362–71
9	Rocha BM, Filos D, Mendes L, Vogiatzis I, Perantoni E, Kaimakamis E, Natsiavas P, Oliveira A, Jacome C, Marques A, Paiva RP, Chaouvarda I, Carvalho P, Maglaveras N. A respiratory sound database for the development of automated classification. In <i>Precision Medicine Powered by pHealth and Connected Health</i> ; 2017, pp. 33–37
10	Jakovljevic N, Loncar-Turukalo T. Hidden Markov model based respiratory sound classification. <i>Precis Med.</i> 2017;17:39–43.
11	Dennis J, Tran HD, Li H. Spectrogram image feature for sound event classification in mismatched conditions. <i>IEEE Signal Process Lett.</i> 2011;18(2):130–3
12	Pan SJ, Yang Q. A survey on transfer learning. <i>IEEE Trans Knowl Data Eng.</i> 2010;22:10