2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Crackle and wheeze detection in lung sound signals using convolutional neural networks

Pedro Faustino, Jorge Oliveira and Miguel Coimbra.

Abstract— **Respiratory diseases are among the leading causes of death worldwide. Preventive measures are essential to avoid and increase the odds of a successful recovery. An important screening tool is pulmonary auscultation, an inexpensive, noninvasive and safe method to assess the mechanics and dynamics of the lungs. On the other hand, it is a difficult task for a human listener since some lung sound events have a spectrum of frequencies outside of the human hearing ability. Thus, computer assisted decision systems might play an important role in the detection of abnormal sounds, such as crackle or wheeze sounds. In this paper, we propose a novel system, which is not only able to detect abnormal lung sound events, but it is also able to classify them. Furthermore, our system was trained and tested using the publicly available ICBHI 2017 challenge dataset, and using the metrics proposed by the challenge, thus making our framework and results easily comparable. Using a Mel Spectrogram as an input feature for our convolutional neural network, our system achieved results in line with the current state of the art, an accuracy of 43%, and a sensitivity of 51%.**

I. INTRODUCTION

According to the World Health Organization (WHO) [1], Chronic Respiratory Diseases (CRDs) are among the leading causes of death in the world. More than 3 million people die each year from Chronic Obstructive Pulmonary Diseases (COPDs), which is approximately 6% of all deaths worldwide. COPD is a non-curable progressive life-threatening lung condition, that restricts lung airflow and predisposes to exacerbations and serious illness, but treatment can relieve symptoms and reduce the risk of death. COPD is not a single disease, but a term used to describe chronic lung diseases that restrict lung airflow. The main causes of COPD are smoking and indoor and outdoor air pollution, other non-avoidable causes such as age and heredity. The most effective method to screen and provide a COPD diagnosis is through a pulmonary auscultation. It is a non-invasive, fast, cheap and easy procedure to assess the state of the patient's lungs [2]. However, the diagnosis process is highly dependent on the physician's experience and ear acuity. There are several lung auscultation spots in the chest, sides and back, with different sound characteristics corresponding to the different lung areas. With the recent development and improvement of digital stethoscopes [3], we now have the ability to digitize lung sound signal. This allows the usage of computer assisted decision systems (CADs) based on lung sound auscultation. Additionally, with the development of wireless services and the Internet of Things (IoT), the benefits of a fully automated diagnosis could be spread worldwide, making them faster and more accessible to the patients, especially when combined with cloud service technology [3]. On the other hand, there are several issues on the analysis of lung sounds, since these are immensely dependant on several factors, such as auscultation stop, patient position, airflow intensity, age, weight, gender, etc [4]. In this paper, we aim to successfully detect and classify adventitious sounds in lung sound signals, through signal processing techniques and deep artificial neural networks, namely convolutional neural network architectures (CNN).

P. Faustino is with the Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre 1021/1055, 4169-007 Porto, Portugal (email: pedr.faustino@gmail.com). J. Oliveira is with Universidade Portucalense Infante D. Henrique, Rua Dr. António Bernardino de Almeida, 541 4200-072 Porto, Portugal (email: jorgefisicomat@gmail.com). M.Coimbra is with INESC TEC, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal and the Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre 1021/1055, 4169-007 Porto, Portugal (email: mcoimbra@dcc.fc.up.pt). This work is financed by national funds through FCT - Fundação para a Ciência e Tecnologia, within the scope of the project DSAIPA / AI / 0083/2020

A. Contribution

The contributions of this work are the following:

- The application of CNNs for the detection and classification of crackles and wheezes.
- The usage of a novel set of lung sound features for the detection of adventitious lung sounds.

This is paper is organized as follow. In Section II, we present a brief description of the state-of-art algorithms on lung sound detection and classification and we provide some background on abnormal lung sounds. In Section III, the signal processing pipeline is described and in Section IV our neural network is explained in detail. In Section V, the experimental setup is defined and results are presented and discussed in Section VI. Finally, in Section VII, conclusions are withdrawn.

II. STATE-OF-ART ALGORITHMS AND BACKGROUND

Since the International Conference on Biomedical and Health Informatics (ICBHI) 2017, released a respiratory sound database, there have been five studies [5], [6], [7], [8], [9], that we are aware of that have used it. The most common algorithms used to detect crackles and wheezes are artificial neural networks (ANN), support vector machines [10], k-nearest neighbors [11] and Gaussian mixture models [12]. The most common ANN are the standard multi-layer perceptron, with the exception of [6] and [5] which respectively use a CNN and recurrent neural network architecture (RNN). In both studies, the system aims to detect abnormal lung sounds. Other approaches used Hidden Markov models to infer the most likely state sequence of events in a lung sound signal. The presence or not of crackles and wheezes in the signal

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

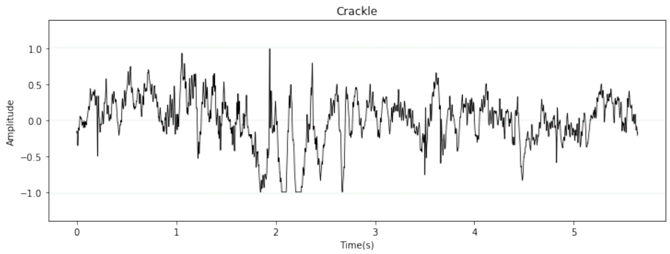

Fig. 1. A crackle sound example.

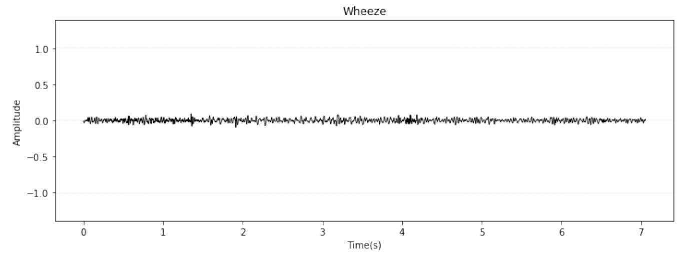

Fig. 2. A wheeze sound example.

is inferred [7] by analyzing the generated state sequences. Finally, Kochetov *et al.* [5] proposed a novel Noise Masking Recurrent Neural Network (NMRNN) to detect abnormal respiratory sounds. They achieved the current state-of-art results a Sensitivity (Se) of 0.56 and a Specificity (Sp) of 0.74 to the aforementioned task.

A. Physiology of Lung Sounds

Abnormal breath sounds are described by the absence or reduced intensity of sounds. Adventitious sounds are additional respiratory sounds superimposed on normal breath sounds. They can be continuous (like wheezes) or discontinuous (such as crackles), and some can be both (like squawks).

- **Crackles** are discontinuous adventitious sounds generated during inspiration (in general). A crackle can be characterized as fine (short duration) or coarse (long duration) [2]. Coarse crackles occur during the beginning of an inhalation and are indicative of a chronic bronchial disease. When in the middle of an inhalation, these are indicative bronchiectasis, and when in the end of an inhalation, these are indicative of a peripheral bronchi. Fine crackles are indicative of a peripheral bronchi. It is generally accepted that the duration of a crackle is lower than 20 ms and the frequency range is between 100 and 200 Hz [2], see Figure 1.
- **Wheeze** is a continuous adventitious sound [2]. Acoustically, it is a periodic wave and lasts more than 100ms. Wheezes are usually associated with an airway obstruction resulted from various causes. The frequency of wheezes lies within 100 and 2500 Hz [2], see Figure 2.

It is important to notice that the detection of crackles and wheezes lies within the range of 100 to 2500 Hz, therefore any other sound outside this range, such as noise, can be safely discarded or filtered without a significant loss of the quality of the adventitious sounds.

III. Signal Processing Pipeline

Since the dataset is composed by lung sounds at different sampling rates, these are resampled to 6000 Hz, using the method proposed by [13]. To remove signal noise artifacts, we applied a 12th order Butterworth bandpass filter [14] with cutoff frequencies of 100 Hz to 2500 Hz. The signal is normalized using a Z-score function. Afterwards, Power Spectrum Density (PSD), Mel Spectrum (MS), Discrete Fourier Transform (DFT) and Mel Frequency Ceptral Coefficients (MFCC) are extracted from the normalized lung sound signal. The PSD and MS are converted to the decibel scale (0 to -80), and further normalized using a min-max normalization. The MFCCs are normalized using a z-score function. Furthermore several sizes of Mel filter banks have been tested, our best results were obtained 64 filters. The window size chose to compute the DFT of the PSD, MS and MFCC is 512 (128 ms).

IV. Deep-Learning Model

In this paper, we implemented a CNN based algorithm due to its innate ability to detect local patterns in a grid-like data structure. A diagram of the proposed one-dimensional network is presented in Figure 3. When lung sound signals are straightforward used, the CNN is composed by 1D convolutional layers. As for the other features (PSD, PS, MFCC), the net is composed by 2D convolutional layers. The CNN is made by 3 convolutional layers, 1 fully-connected layer and a Global Max Pooling (GMP) layer between the final convolutional layer and the fully connected layer in order to transform the dynamic sized grid into a static size feature vector.

In the last layer, the most likely class is inferred. To do so a softmax activation function is used. We use Leaky ReLU with an alpha of 0.001 as the activation function for the convolutional layers. We used the unit norm constraint for the weights of the individual kernels of the convolutional layers and the fully-connected layer. For a CNN made by 2D convolutional layers, we use 64 kernels of size (3,3) and stride of (1,1) on all layers. The receptive field on the first layer is 256 ms and 3 Hz. In the second layer is 384 ms and 5 Hz. In last layer is 512 ms and 7 Hz.
For a CNN made by 1D convolutional layers, we use 16 kernels of size 3 and stride of 1 on all layers. The CNN made by 2D convolutional layers has 74,756 trainable parameters, on the other hand the CNN made by 1D convolutional layers has 63,428 trainable parameters. Increasing the number of convolutional layers, on both nets lead to overfitting behavior. On the other hand, fewer layers damaged the model's ability to learn the signal patterns.

V. Experimental Setup

A. Materials

The ICBHI 2017 respiratory sound database [15] was part of a scientific challenge aimed to test and compare

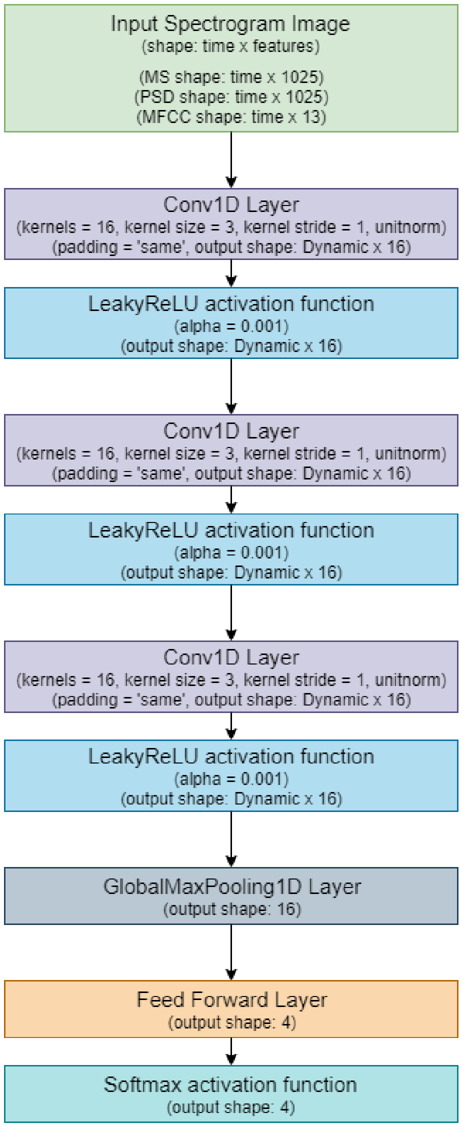

Fig. 3. A diagram of the proposed CNN model.

the robustness of state-of-the-art techniques on lung sound analysis. The goal was to classify each individual respiratory cycle into one of four classes: Normal, crackle, wheeze, both.

The dataset consists of a set of respiratory sound recordings and their corresponding annotation files. The audio samples were collected independently by two research teams, the respiratory research and rehabilitation laboratory of the school of health sciences, University of Aveiro, Portugal (Lab3R) and the Aristotle University of Thessaloniki, Greece (AUTH). The dataset contains 920 annotated audio recordings collected from 126 participants. Each audio recording was obtained using a multi-channel or single-channel acquisition system on several auscultation spots. The auscultation spots are: Anterior left (Al), Anterior right (Ar), Lateral left (Ll), Lateral right (Lr), Posterior left (Pl), Posterior right (Pr) and Trachea (Tc). Each recording was manually annotated over several respiratory cycles. On each cycle, the starting, ending timestamp and the presence of crackles or wheezes was annotated by health professionals.

B. Metrics of Performance

In [15], a set of metrics was proposed: Sp, SE, Average score (AS) and Harmonic score (HS). These metrics are calculated as follows:

$$SE = (Cc + Ww + Bb)/(C + W + B) \quad (1)$$

$$SP = Nn/N \quad (2)$$

$$AS = (SE + SP)/2 \quad (3)$$

$$HS = \frac{(2 \times SE \times SP)}{(SE + SP)} \quad (4)$$

where N is the number of normal sounds, Nn is the number of correctly classified normal sounds, C is the number of crackle sounds, Cc is the number of correctly classified crackle sounds, W is the number of wheeze sounds, Ww is the number of correctly classified wheezes, B is the number of sounds that contain both crackle and wheeze sounds and Bb is the number of correctly classified sounds that contain both adventitious sounds. The dataset was split into training (60%) and testing (40%) sets, 2063 respiratory cycles from 79 participants were included in the training set, while 1579 respiration cycles from 49 patients were included in the testing set. Recordings from the same patient are only used for training or testing purposes, thus avoiding biasing in our predictions. Due to the class imbalance, the training is performed in the following way, for each epoch:

- Sample a random number of samples from each class.
- Shuffle samples.
- Train the model using mini-batches of size one.

The maximum number of samples of each class is equal to the number of samples of the minority class. Under sampling is applied to both train and test sets during the training step. Lastly, five-fold cross validation [16] was used to measure the robustness of proposed algorithms.

The weights of the CNN are learned using a stochastic gradient descent optimization method [17]. The learning rate is set to 0.01 over the first 180 epochs, 0.001 for the next 60 epochs and 0.001 for the last 60 epochs.

VI. Results and Discussion

Training CNN models in the current dataset is indeed a great challenge: the dataset is unbalanced, a high variability on the respiratory cycles, unbalance records per auscultation spot, different sampling rates, different equipment properties, noise, artifacts, etc. Nevertheless, the average scores over the five folds are presented in Table I.

The results in Table I show that MS features are the most generic. They detect adventitious sounds more accurately than any other feature, but also the worst SP. The weakness of the straightforward usage of lung sounds is evident, whereas it can easily distinguish normal from crackle sounds, it fails to detect wheeze sounds. It is worthy to notice that, although MFCC features are standard, they did not generalize as

	Net size	Acc	AS	HS	Sp	SE	"Normal" Recall	"crackle" Recall	"wheeze" Recall	"both" Recall
Lung Sounds	63,428	37%	0.37	0.36	0.41	0.33	0.41	0.45	0.22	0.09
PSD	74,756	40%	0.40	0.39	0.37	0.42	0.37	0.52	0.33	0.28
MS	74,756	**43%**	**0.43**	**0.42**	0.36	**0.51**	0.36	**0.62**	**0.37**	**0.34**
MFCC	74,756	43%	0.42	0.42	**0.42**	0.42	**0.42**	0.55	0.26	0.26

TABLE I

RESULTS FROM THE USAGE OF LUNG SOUNDS, PSD, MS AND MFCC FEATURES IN A CNN MODEL. THE BEST RESULTS FOR EACH METRIC ARE HIGHLIGHTED IN BOLD.

well as it was expected. On the other hand, MS features consistently had the smallest gap between the train/test loss and train/test accuracy, followed by PSD, MFCC and the lung sound signals. We hypothesize that MS is the most advisable feature to the detection of crackles and wheeze in lung sounds, because the transformation to the Mel space might mimics the human ear and it creates a "smoothing" effect on the frequency component of the signal.

VII. CONCLUSIONS

In this paper, a set of features are feed into a CNN to detect pulmonary diseases. We present different functions to convert lung sound signals into a 2D image for classification purposes. We also experimented the usage of a 1D CNN for the processing and classification of the lung sound signal. We found that MS when feed into CNN achieves results in line with the current state-of-art [5]. For future work, we intend test and design new deep learning architectures, based on more sophisticated deep neural net configurations.

REFERENCES

[1] J. Bousquet and N. Kaltaev, "Global surveillance, prevention and control of chronic respiratory diseases : a comprehensive approach / edited by jean bousquet and nikolai khaltaev," pp. vii, 146 p., 2007.

[2] M. Sarkar, I. Madabhavi, N. Niranjan, and M. Dogra, "Auscultation of the respiratory system," *Annals of thoracic medicine*, vol. 10, pp. 158–168, 07 2015.

[3] S. Swarup and A. N. Makaryus, "Digital stethoscope: technology update," *Medical devices (Auckland, N.Z.)*, vol. 11, p. 29—36, 2018. [Online]. Available: https://europepmc.org/articles/PMC5757962

[4] S. Reichert, R. Gass, C. Brandt, and E. Andrès, "Analysis of respiratory sounds: State of the art," *Clinical medicine. Circulatory, respiratory and pulmonary medicine*, vol. 2, p. CCRPM.S530, 2008, pMID: 21157521. [Online]. Available: https://doi.org/10.4137/CCRPM.S530

[5] K. Kochetov, E. Putin, M. Balashov, A. Filchenkov, and A. Shalyto, *Noise Masking Recurrent Neural Network for Respiratory Sound Classification: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4–7, 2018, Proceedings, Part III*, 10 2018, pp. 208–217.

[6] D. Perna, "Convolutional neural networks learning from respiratory data," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2018, pp. 2109–2113.

[7] N. Jakovljević and T. Lončar-Turukalo, "Hidden markov model based respiratory sound classification," in *Precision Medicine Powered by pHealth and Connected Health*, N. Maglaveras, I. Chouvarda, and P. de Carvalho, Eds. Singapore: Springer Singapore, 2018, pp. 39–43.

[8] A. Rao, S. Chu, N. Batlivala, S. Zetumer, and S. Roy, "Improved detection of lung fluid with standardized acoustic stimulation of the chest," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 6, pp. 1–7, 2018.

[9] G. Chambres, P. Hanna, and M. Desainte-Catherine, "Automatic detection of patient with respiratory diseases using lung sound analysis," in *2018 International Conference on Content-Based Multimedia Indexing (CBMI)*, Sep. 2018, pp. 1–6.

[10] M. M. Azmy, "Classification of lung sounds based on linear prediction cepstral coefficients and support vector machine," in *2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, Nov 2015, pp. 1–5.

[11] J. H. L. Serato and R. Reyes, "Automated lung auscultation identification for mobile health systems using machine learning," in *2018 IEEE International Conference on Applied System Invention (ICASI)*, April 2018, pp. 287–290.

[12] J. Chien, H. Wu, F. Chong, and C. Li, "Wheeze detection using cepstral analysis in gaussian mixture models," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug 2007, pp. 3168–3171.

[13] W. Putnam and J. Smith, "Design of fractional delay filters using convex optimization," in *Proceedings of 1997 Workshop on Applications of Signal Processing to Audio and Acoustics*, 1997, pp. 4 pp.–.

[14] S. Butterworth, "On the theory of filter amplifiers. the wireless engineer," *From 1923 to 1930, the journal was called Experimental Wireless and the Radio Engineer*, vol. 7, pp. 536–541, 1930. [Online]. Available: https://ci.nii.ac.jp/naid/10030057017/en/

[15] B. M. Rocha, D. Filos, L. Mendes, G. Serbes, S. Ulukaya, Y. P. Kahya, N. Jakovljević, T. L. Turukalo, I. M. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jácome, A. Marques, N. Maglaveras, R. P. Paiva, I. Chouvarda, and P. de Carvalho, "An open access database for the evaluation of respiratory sound classification algorithms." *Physiological measurement*, vol. 40 3, p. 035001, 2019.

[16] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-validation." in *Encyclopedia of Database Systems*, L. Liu and M. T. Özsu, Eds. Springer US, 2009, pp. 532–538. [Online]. Available: http://dblp.uni-trier.de/db/reference/db/c.html#RefaeilzadehTL09

[17] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.