











Detection of squawks in respiratory sounds of mechanically ventilated COVID-19 patients

Bruno Machado Rocha¹ , Diogo Pessoa¹ , Grigorios-Aris Cheimariotis² ,
Evangelos Kaimakamis^{2,3} , Serafeim-Chrysovalantis Kotoulas³ , Myrto Tzimou³ ,
Nicos Maglaveras² , Alda Marques⁴ , Paulo de Carvalho¹  and Rui Pedro Paiva¹ 

Abstract—Mechanically ventilated patients typically exhibit abnormal respiratory sounds. Squawks are short inspiratory adventitious sounds that may occur in patients with pneumonia, such as COVID-19 patients. In this work we devised a method for squawk detection in mechanically ventilated patients by developing algorithms for respiratory cycle estimation, squawk candidate identification, feature extraction, and clustering. The best classifier reached an F1 of 0.48 at the sound file level and an F1 of 0.66 at the recording session level. These preliminary results are promising, as they were obtained in noisy environments. This method will give health professionals a new feature to assess the potential deterioration of critically ill patients.

Index Terms—Respiratory Sounds, Audio Signal Processing, Intensive Care

I. INTRODUCTION

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). While most patients who develop symptoms recover from the disease without needing hospital treatment, about 15% become seriously ill and require oxygen and 5% become critically ill and need intensive care [13]. In the intensive care unit (ICU), health professionals use various means to monitor patients, including chest x-rays, computed tomography scans, or auscultation of respiratory sounds.

Respiratory sounds are a noninvasive and objective marker to assess patients' respiratory condition [7]. Adventitious sounds are additional respiratory sounds superimposed on normal breath sounds [20]. They are mainly composed by continuous (wheezes) or discontinuous (crackles) sounds [12].

Squawks are short inspiratory adventitious sounds, containing sinusoidal and noisy components, with a duration of 50 to 400 ms [3], [4]. A squawk usually

appears as a short wheeze preceded by a crackle [15]. The typical fundamental frequency of the sinusoidal component of a squawk is between 200 and 300 Hz [3]. Squawks are related to certain restrictive (like allergic alveolitis and diffuse interstitial fibrosis) or acute (like severe pneumonia and bronchiolitis obliterans) lung pathologies [19]. When they appear in acute respiratory distress syndrome (ARDS) critically ill patients in the ICU they can signal either a premature progression to fibrotic processes of ARDS or the onset of a localized ventilator-associated pneumonia. Nevertheless, little attention has been paid to this adventitious sound, and further education on its identification and interpretation has been recommended [12].

Several methods have been devised to detect squawks. Lacunarity was used to discriminate between fine crackles, coarse crackles, and squawks, achieving an accuracy of 100% [5]. Neural networks have also been used to classify squawks along with other classes of adventitious sounds [9], [2], reaching accuracies higher than 90%. However, all those works used very small datasets containing less than 10 squawks. To the best of our knowledge, squawk detection in an ICU setting has not been performed before.

In this work, we developed a method for the detection of squawks in respiratory sounds of mechanically ventilated patients with COVID-19. Our method encompasses several steps: respiratory cycle estimation, identification of possible squawks, feature extraction, and clustering.

II. MATERIALS AND METHODS

A. Ethics and participants

Data used in this study were collected in the ICU of "G. Papanikolaou" General Hospital, Thessaloniki, Greece. The study was approved by the Ethics Committee of the same hospital. Informed consent was not possible to obtain, due to the restrictions to visits to the hospital posed by the current pandemic and was deemed acceptable practice by the ethics committee of the institution.

B. Data collection and annotation

259 sound files (SF) (duration: 15 s) were acquired from 2 chest locations (posterior basal left and right) of 29 (10 women, 19 men) mechanically ventilated (pressure controlled and volume controlled ventilation) COVID-19 patients, with an average age of 65.4 ± 11.4 years. The

¹University of Coimbra, Centre for Informatics and Systems of the University of Coimbra, Department of Informatics Engineering, 3030-290 Coimbra, Portugal. {bmrocha, dpessoa, carvalho, ruipedro}@dei.uc.pt

²Laboratory of Computing, Medical Informatics and Biomedical Imaging Technologies, Aristotle University of Thessaloniki, Thessaloniki, 54636, Greece. {ncheimar, nicmag}@auth.gr

³1st Intensive Care Unit, "G. Papanikolaou" General Hospital, Thessaloniki, Greece. vkaimak@med.auth.gr, akiskotoulas@hotmail.com, tzimzapatista@gmail.com

⁴Lab3R - Respiratory Research and Rehabilitation Laboratory, School of Health Sciences (ESSUA), and Institute of Biomedicine (iBiMED), University of Aveiro, Aveiro, Portugal. amarques@ua.pt

acquisition sensor was the Littmann 3200 electronic stethoscope (3M, St. Paul, Minnesota, USA), which produced audio files with a sampling frequency of 4 kHz.

Data collection of each patient was divided by recording sessions (RS), where one of three physicians acquired at least two recordings and noted the respiratory rate (RR) at the time. A total of 123 RS were carried out.

Manual annotation of respiratory sounds is a very time-consuming task requiring expert knowledge from health professionals. Given the time limited availability, due to the current phase of the pandemic, only one experienced physician annotated the presence/absence of squawks in each SF. The annotation was blind, i.e., the annotator did not know to which patient or RS did a SF belong. Furthermore, the annotations were purely aural, an aspect that should be addressed in future work by also providing visual information to the annotators (e.g., time-expanded waveforms and spectrograms). Posterior basal locations were selected because they maximize lung sound intensity [14] while minimizing heart sound interference.

C. Respiratory cycle estimation

Respiratory cycle (RC) estimation is a complex problem, especially when subjects with respiratory diseases are involved, since breathing pattern and lung sounds are known to change in the presence of respiratory diseases [8]. A good RC estimation is important for squawk detection, given that squawks occur exclusively in the inspiration phase [3]. However, RC estimation is easier when the RR is known and fixed, because one well-estimated inspiration onset is enough to find out all inspiration onsets in a SF. First, we applied a fourth-order elliptic band-pass filter between 100 and 400 Hz using the MIR Toolbox [10]. Then, we computed the RMS energy in 100 ms frames with 90% overlap. Next, we performed peak-picking on the RMS curve and we calculated the distance between each peak, creating a distance matrix. As all patients in this study were mechanically ventilated, we could assume that the RR was fixed and that all the RCs in the same RS had the same duration. Then, we computed the remainder after division (RAD) between each peak and RC duration. Assuming a typical inspiration:expiration ratio of 1:2, we found the peaks with a RAD of less than the duration of half an inspiration. The mode of the resulting vector should correspond to possible inspiratory peaks, assuming that the highest of the remaining energy peaks should happen during an inspiration. We searched for the last frame in the previous second where the RMS value crosses a given threshold to find the beginning of the main inspiration. The following equations show the formulas for estimating the RC duration and for discovering the beginning of the main inspiration.

$$RCDuration = 60/RR \quad (1)$$

$$InspirationDuration = RCDuration/3 \quad (2)$$

$$Threshold = HighestPeakValue/4 \quad (3)$$

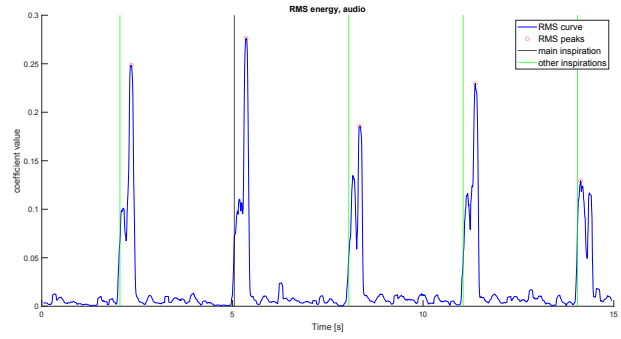


Fig. 1: Respiratory cycle estimation using the RMS energy.

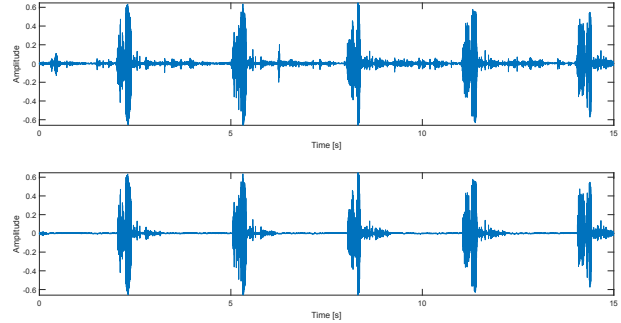


Fig. 2: Audio waveforms before and after removing non-relevant parts.

Finally, by subtracting and adding RCD to the estimated beginning of the main inspiration, we got a vector containing all the inspiration onsets. An example of this process can be seen in Figure 1.

D. Identification of possible squawks

Having a vector of estimated inspiration onsets and knowing that squawks are inspiratory sounds, we can define time intervals where squawks are most likely to occur. We established that each squawk interval started at the inspiration onset and had the duration of one inspiration. Then, we added attenuated pink noise to the signal and removed the non-relevant intervals of the audio signal, as shown in Figure 2. The addition of pink noise was inspired by the ensemble empirical mode decomposition (EEMD) approach, which consists of sifting an ensemble of white noise-added signal and considering the mean as the final result [21].

The next step was to decompose the signal into intrinsic mode functions (IMFs) using EMD [6]. Rilling and Flandrin summarize the EMD rationale by the motto “signal = fast oscillations superimposed to slow oscillations”, with iteration on the slow oscillations considered as a new signal [17]. Assuming the sinusoidal part of a squawk is a high-frequency narrow-band signal, we extracted just the first IMF (IMF1). Then, we obtained a Bump scalogram by computing the continuous wavelet transform between 125 and 1000 Hz. The wavelet magnitude was then normalized by its maximum value and a binary image is generated by applying a magnitude

threshold, as shown in the following equation:

$$signal = \begin{cases} 1, & \text{magnitude} > \text{threshold} \\ 0, & \text{magnitude} \leq \text{threshold} \end{cases} \quad (4)$$

Three threshold values were tested: 5%, 25%, and 50%. Figure 3 plots the resulting binary images after applying different thresholds.

Subsequently, we selected all non-flat connected components (CC) that have durations between 25 ms and 400 ms. These values were chosen considering that the sinusoidal component encompasses at least half of each squawk's duration.

E. False positive elimination

Before extracting features for each CC, we ascertained whether a file had certain characteristics which would make it unlikely to contain squawks. To that effect, we computed the enhanced magnitude spectrum of the autocorrelation of IMF1 using the spectral product method with compression factor $M = 2$ [1]. If the highest peak of the spectrum was below 75 Hz (half of a fundamental frequency of 150 Hz) or above 500 Hz, the file was discarded. In addition, we computed the spectral flatness in 500 ms frames with 50% overlap and discard files whose minimum flatness is below 50%, a loose threshold given that most squawks should have a minimum flatness well below 50%.

F. Feature extraction

In this step, we extracted 17 features for each event, described in Table I. Spectral features were extracted from the spectrum of IMF1 at each event, while the other features were estimated from the binary image of each CC.

TABLE I: Description of the 17 extracted features.

Feature	Description
Duration	Duration of event
Fundamental Frequency	Minimum frequency
Frequency Range	Frequency range
Zero-Crossing Rate	Number of zero-crossings per second
IMF1 Peaks	Number of peaks above 1/4 of the maximum amplitude of IMF1 of each event
Graphical Extent	Ratio of pixels in the CC to pixels in the total bounding box
Graphical Perimeter Area	Ratio of pixels around the boundary of the CC to pixels in the CC
Spectral Centroid	Center of mass of the spectral distribution
Spectral Crest	Ratio between the maximum spectral value and the arithmetic mean of the energy spectrum value [16]
Spectral Entropy	Estimation of the complexity of the spectrum
Spectral Flatness	Estimation of the noisiness of a spectrum
Spectral Kurtosis	Measure of the flatness of a distribution around its mean value
Spectral Rolloff	Frequency such that 95% of the total energy is contained below it
Spectral Skewness	Measure of the asymmetry of a distribution around its mean value
Spectral Slope	Linear regression of the magnitude spectrum
Spectral Spread	Variance of the spectral distribution [11]
Harmonic Ratio	Maximum of the normalized autocorrelation

G. Clustering

After centering the data to have median 0, we partitioned the observations into clusters using the k-medoids algorithm. To determine the number of clusters, we test $k = \{2, 3, \dots, n\}$, where n is the number of squawk intervals, and choose the value of k that maximizes the median of the silhouette, which is a measure of how similar each point is to points in its own cluster, when compared to points in other clusters [18]. Subsequently, we had to choose the cluster which was most likely to contain squawks. Considering that squawks are usually repeatable and tend to appear in every inspiration, they should form a consistent cluster, with each point having similar characteristics. A composite of 9 features was then used to choose the best candidate cluster. More specifically, we assumed that:

- Spectral crest, kurtosis, skewness, and harmonic ratio should be high;
- Spectral entropy, flatness, rolloff, slope, and spread should be low.

Therefore, the medoid that best approximated the ideal cluster was chosen. Next, assuming that there could be only one squawk per inspiration, we chose for each inspiration the squawk whose fundamental frequency was closer to the median fundamental frequency of the cluster. Finally, we discarded squawks that did not conform to some conservative rules:

- Fundamental frequency and spectral centroid should be lower than 500 Hz;
- Spectral skewness and frequency range should be positive;
- There should be at least 10 IMF1 Peaks, corresponding to a stable signal of at least 100 Hz.

If at least two squawks survived after this process, the file was classified as containing squawks.

H. Evaluation Measures

We used the following measures to evaluate the performance of the algorithms:

$$Precision = \frac{TP}{(TP + FP)} \quad (5)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (6)$$

$$F1Score(F1) = \frac{(2 \times Precision \times Recall)}{(Precision + Recall)} \quad (7)$$

$$MatthewsCorrCoef(MCC) = \frac{((TP \times TN) - (FP \times FN))}{\sqrt{((TP + FP)(TP + FN)(TN + FP)(TN + FN))}} \quad (8)$$

where TP (True Positives) are files with squawks that are correctly classified; TN (True Negatives) are files with no squawks that are correctly classified; FP (False Positives) are files that are incorrectly classified as containing squawks; FN (False Negatives) are files with squawks that are incorrectly classified.

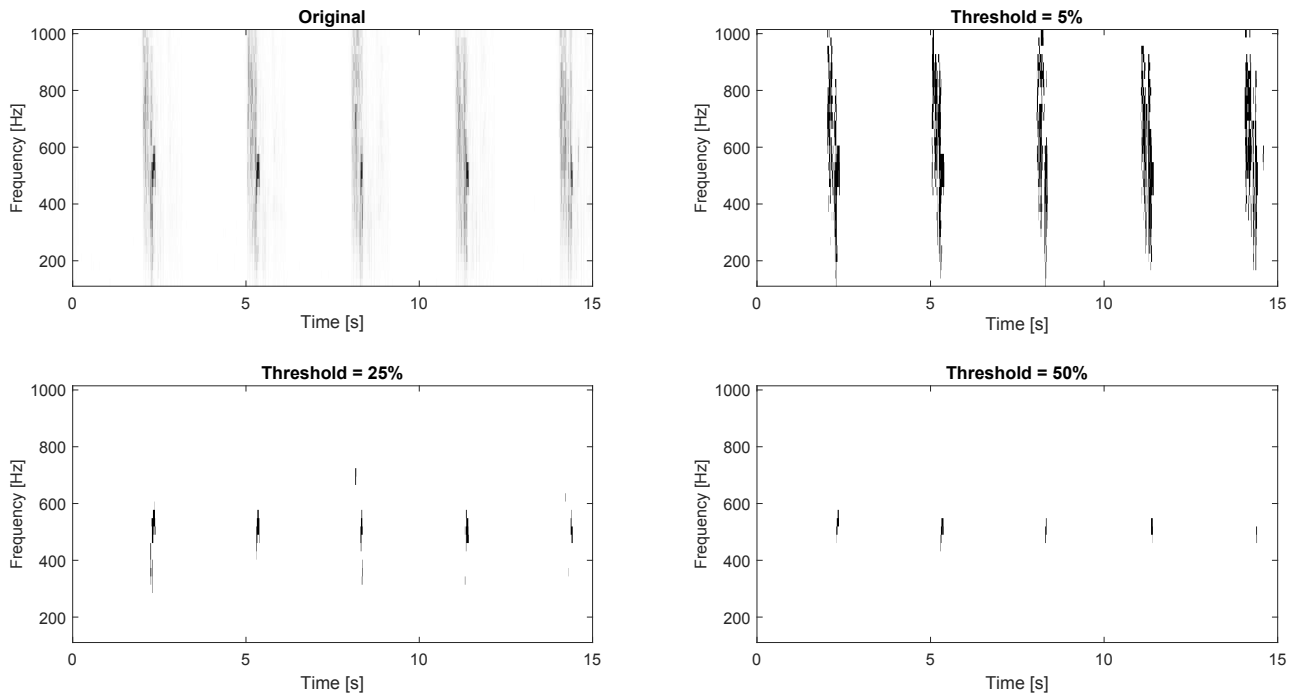


Fig. 3: Binary images after applying different thresholds. The higher the threshold, the sparser the resulting image.

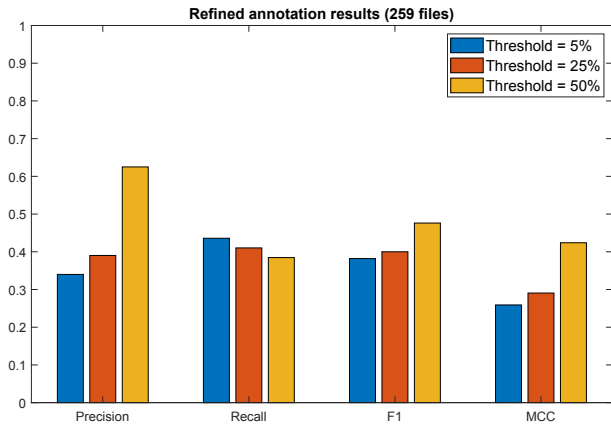


Fig. 4: Results at the SF level.

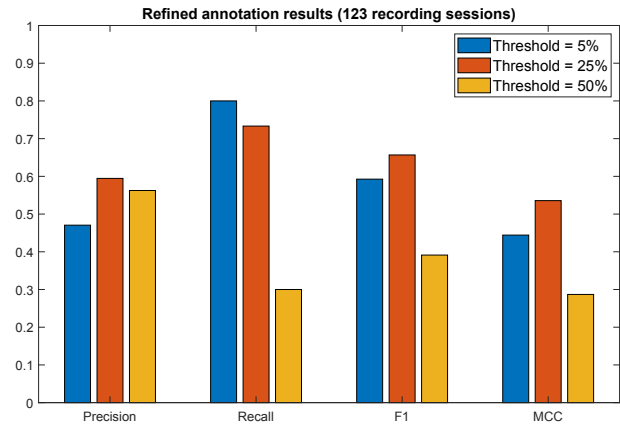


Fig. 5: Results at the RS level.

III. EVALUATION

In this section, we analyze the performance of the squawk detection method in two ways: at the SF level and at the RS level.

At the SF level, the algorithm with the best performance had a threshold of 50%, attaining a F1 of 0.48 and a MCC of 0.42. Figure 4 shows the results at the SF level. At the RS level, the algorithm with 25% threshold achieved the best results, reaching a F1 of 0.66 and a MCC of 0.54, as displayed in Figure 5.

As can be seen in Figure 4 and Figure 5, the trade-off between precision and recall is mediated by the wavelet threshold, which behaves as a dial that health professionals can use to decide if the detection should be looser or stricter. The results at the RS level are the most meaningful

in practical terms, as the main output that health professionals expect is to know if squawks are detected in a given RS and to compare those results to previous and subsequent RS.

Considering just the 25% threshold algorithm, another noteworthy fact was that, while the number of FP at the SF level was 25, it was only 13 at the RS level. Furthermore, only 3 FP were from patients that presented no squawks during their stay in the ICU, one from each. The other 10 FP were from RS preceding RS with annotated squawks. Thus, we can speculate that the algorithm was sensitive to squawks before they were audible by the expert. A larger dataset would be necessary to test that hypothesis. Regarding FN, there were 23 at the SF level and 8 at the RS level, with only 2 files from 2 patients that had squawks. Figure 6 shows examples of a TP, a FP and a FN.

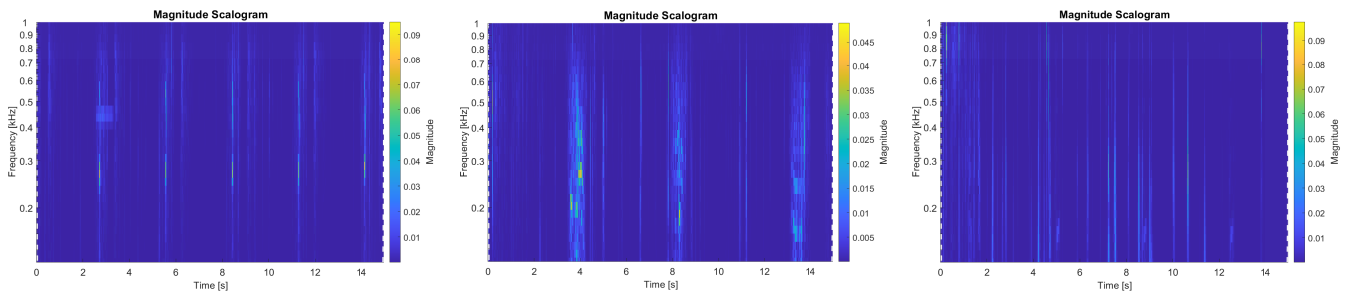


Fig. 6: Examples of true positive (left), false positive, (center), and false negative (right).

The TP example corresponds to a prototypical case of a file with squawks: the squawks are repeatable, i.e., they appear in every inspiration, with a fundamental frequency around 250 Hz and a duration of 50 ms. The components that can be seen in the FP example are probably inspiratory wheezes, as annotated by the health professionals in the rapid annotations. Regarding the FN example, it demonstrates how a robust RC estimation is needed to improve the results, as the squawk components visible in the scalogram are outside the boundaries of the intervals automatically defined as relevant. Thorough analysis of RC estimation was outside the scope of this paper and should be addressed in future work.

IV. CONCLUSION

In this work, we proposed an algorithm for the detection of squawks in ICU patients and analyzed how the wavelet threshold for creating the binary images affected the performance at the SF level and at the RS level. This method will allow physicians to be alerted about the occurrence of squawks, helping them to timely assess the potential deterioration of a critically ill ARDS patient.

ACKNOWLEDGMENT

This work was partially supported by: Fundação para a Ciência e Tecnologia (FCT) PhD scholarships SFRH/BD/135686/2018 and DFA/BD/4927/2020; WELMO project, co-funded by the Horizon 2020 Framework Programme of the European Union under grant agreement 825572; FCT project Lung@ICU under grant reference DSAIPA/AI/0113/2020; Fundo Europeu de Desenvolvimento Regional (FEDER) through Programa Operacional Competitividade e Internacionalização (COMPETE) and FCT under the project UIDB/04501/2020 and POCI-01-0145-FEDER-007628—iBiMED.

REFERENCES

[1] M. Alonso, B. David, and G. Richard. A study of tempo tracking algorithms from polyphonic music signals. In *4 COST 276Workshop*, pages 1–5, 2003.

[2] D. Bardou, K. Zhang, and S. M. Ahmad. Lung sounds classification using convolutional neural networks. *Artificial Intelligence in Medicine*, 88:58–69, 2018.

[3] A. Bohadana, G. Izbicki, and S. S. Kraman. Fundamentals of Lung Auscultation. *New England Journal of Medicine*, 370(8):744–751, February 2014.

[4] J. E. Earis, K. Marsh, M. G. Pearson, and C. M. Ogilvie. The inspiratory "squawk" in extrinsic allergic alveolitis and other pulmonary fibroses. *Thorax*, 37(12):923–6, 1982.

[5] L. J. Hadjileontiadis. A texture-based classification of crackles and squawks using lacunarity. *IEEE Transactions on Biomedical Engineering*, 56(3):718–732, 2009.

[6] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Snin, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995, 1998.

[7] C. Jácome and A. Marques. Computerized Respiratory Sounds in Patients with COPD: A Systematic Review. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, 12(1):104–112, January 2015.

[8] C. Jácome, J. Ravn, E. Holsbø, J. C. Aviles-Solis, H. Melbye, and L. A. Bongo. Convolutional neural network for breathing phase detection in lung sounds. *Sensors (Switzerland)*, 19(8):1–10, 2019.

[9] A. Kandaswamy, C. S. Kumar, R. P. Ramanathan, S. Jayaraman, and N. Malmurugan. Neural classification of lung sounds using wavelet coefficients. *Computers in Biology and Medicine*, 34:523–537, 2004.

[10] O. Lartillot and P. Toivainen. Mir in matlab (II): A toolbox for musical feature extraction from audio. *Proceedings of the 8th International Conference on Music Information Retrieval, ISMIR 2007*, pages 127–130, 2007.

[11] A. Lerch. *An introduction to audio content analysis: Applications in signal processing and music informatics*. Wiley-IEEE Press, 2012.

[12] A. Marques and A. Oliveira. *Normal Versus Adventitious Respiratory Sounds*, chapter 10, pages 181–206. Springer International Publishing, Cham, 2018.

[13] W. H. Organization. Coronavirus disease (COVID-19), 2020.

[14] H. Pasterkamp, S. S. Kraman, P. D. DeFrain, and G. R. Wodicka. Measurement of respiratory acoustical signals: Comparison of sensors. *Chest*, 104(5):1518–1525, 1993.

[15] H. Pasterkamp, S. S. Kraman, and G. R. Wodicka. Respiratory Sounds Advances Beyond the Stethoscope. *American Journal of Respiratory and Critical Care Medicine*, 1997.

[16] G. Peeters. A large set of audio features for sound description (similarity and classification) in the CUIDADO project. Technical report, IRCAM, 2004.

[17] G. Rilling and P. Flandrin. One or Two Frequencies? The Empirical Mode Decomposition Answers. *IEEE Transactions on Signal Processing*, 56(1):85–95, 2008.

[18] P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(C):53–65, 1987.

[19] A. Sovijärvi, L. P. Malmberg, G. Charbonneau, J. Vanderschoot, F. Dalmaso, C. Sacco, M. Rossi, and J. E. Earis. Characteristics of breath sounds and adventitious respiratory sounds. *Eur Respir Rev*, 10(77):591–596, 2000.

[20] A. R. A. Sovijärvi, F. Dalmaso, J. Vanderschoot, L. P. Malmberg, and G. Righini. Definition of terms for applications of respiratory sounds. *Eur Respiratory Rev*, 10(77):597–610, 2000.

[21] Z. Wu and N. E. Huang. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 1(1):1–41, 2009.