Automatic Wheeze Segmentation Using Harmonic-Percussive Source Separation and Empirical Mode Decomposition

Bruno Machado Rocha, Diogo Pessoa, Alda Marques, Paulo de Carvalho, and Rui Pedro Paiva

***Abstract*—Wheezes are adventitious respiratory sounds commonly present in patients with respiratory conditions. The presence of wheezes and their time location are relevant for clinical reasons, such as understanding the degree of bronchial obstruction. Conventional auscultation is usually employed to analyze wheezes, but remote monitoring has become a pressing need during recent years. Automatic respiratory sound analysis is required to reliably perform remote auscultation. In this work we propose a method for wheeze segmentation. Our method starts by decomposing a given audio excerpt into intrinsic mode frequencies using empirical mode decomposition. Then, we apply harmonic-percussive source separation to the resulting audio tracks and get harmonic-enhanced spectrograms, which are processed to obtain harmonic masks. Subsequently, a series of empirically derived rules are applied to find wheeze candidates. Finally, the candidates stemming from the different audio tracks are merged and median filtered. In the evaluation stage, we compare our method to three baselines on the ICBHI 2017 Respiratory Sound Database, a challenging dataset containing various noise sources and background sounds. Using the full dataset, our method outperforms the baselines, achieving an F1 of 41.9%. Our method's performance is also better than the baselines across several stratified results focusing on five variables: recording equipment, age, sex, body-mass index, and diagnosis. We conclude that wheeze segmentation has not been solved for real life scenario applications. Adaptation of existing systems to demographic characteristics might be a promising step in the direction of algorithm personalization, which would make automatic wheeze segmentation clinically viable.**

***Index Terms*—Respiratory sound analysis, expert systems, harmonic-percussive source separation, empirical mode decomposition, sound event detection.**

I. Introduction

RESPIRATORY diseases were highly neglected up until 20 years ago as it was believed that nothing could be done except persuading a person to quit smoking and take inhaled medication. These diseases received less funding and public attention than other diseases (cardiovascular, cancer, Alzheimer) [1]. Currently, respiratory diseases are leading causes of morbidity and mortality worldwide [2]. Pulmonary auscultation using a stethoscope is commonly performed to assess the respiratory condition and its clinical usefulness has been increased with the advent of computer-assisted techniques [3]. Most research in this topic has focused on early diagnosis and monitoring [4], but remote monitoring has become a pressing need with the advent of the COVID-19 pandemic [5]. Therefore, automated methods for the analysis of respiratory sounds are increasingly needed to reliably carry out remote auscultation.

Normal respiratory sounds are produced from breathing and heard over the trachea and chest wall, while adventitious respiratory sounds are abnormal sounds that are superimposed on normal respiratory sounds [6]. Adventitious respiratory sounds can be continuous or discontinuous [7]. Henceforth, we will adopt the terminology defined by the standardization of lung sound nomenclature taskforce [8], i.e., continuous adventitious respiratory sounds will be referred to as wheezes, while discontinuous adventitious respiratory sounds will be called crackles.

Crackles are explosive, short, and nonmusical adventitious respiratory sounds that are attributed to the sudden opening and closing of abnormally closed airways [9]. In contrast, wheezes are musical respiratory sounds usually longer than 100 ms and with frequencies ranging from 100 Hz to 1000 Hz, with harmonics that occasionally exceed 1000 Hz [3]. Wheezes are associated with flow limitation and they can be produced by all mechanisms that reduce airway caliber. Clinically, they can be defined by their

Manuscript received 12 May 2022; revised 20 November 2022; accepted 14 February 2023. Date of publication 23 February 2023; date of current version 5 April 2023. This work was supported in part by the FCT - Foundation for Science and Technology, in part by the I.P./MCTES through national funds (PIDDAC), within the scope of CISUC R&D Unit under Projects UIDB/00326/2020 and UIDP/00326/2020, in part by the Ph.D. scholarships under Projects SFRH/BD/135686/2018 and DFA/BD/4927/2020, in part by the WELMO, in part by the Horizon 2020 Framework Programme of the European Union under Grant 825572, in part by the FCT project Lung@ICU under Grant DSAIPA/AI/0113/2020, in part by the Fundo Europeu de Desenvolvimento Regional (FEDER), in part by the Programa Operacional Competitividade e Internacionalização (COMPETE), in part by the iBiMED under Project POCI-01-0145-FEDER-007628, and in part by the FCT under Project UIDB/04501/2020. *(Corresponding author: Bruno Machado Rocha.)*

Bruno Machado Rocha, Diogo Pessoa, Paulo de Carvalho, and Rui Pedro Paiva are with the Department of Informatics Engineering, University of Coimbra, Centre for Informatics and Systems of the University of Coimbra, 3030-290 Coimbra, Portugal (e-mail: bmrocha@dei.uc.pt; dpessoa@dei.uc.pt; carvalho@dei.uc.pt; ruipedro@dei.uc.pt).

Alda Marques is with the Lab3R — Respiratory Research and Rehabilitation Laboratory, School of Health Sciences (ESSUA), and the Institute of Biomedicine (iBiMED), University of Aveiro, 3810-193 Aveiro, Portugal (e-mail: amarques@ua.pt).

This article has supplementary downloadable material available at https://doi.org/10.1109/JBHI.2023.3248265, provided by the authors.

Digital Object Identifier 10.1109/JBHI.2023.3248265

frequency (mono- or polyphonic), intensity, number, duration, and position in the respiratory cycle (inspiratory or expiratory), gravity influence, and respiratory maneuvers [9]. Wheezes have been used for diagnostic purposes in several respiratory conditions, such as asthma [9].

In this work we focus on wheezes and we propose a method for wheeze segmentation. In realistic settings, an algorithm needs to determine the location of wheezes in long recordings. For example, figuring out the proportion of the respiratory cycle occupied by wheezes may be relevant to ascertain the degree of bronchial obstruction [9]. In that scenario, wheeze segmentation is more important than the most common task related to wheeze analysis, wheeze classification, where predetermined sound events are classified as wheezes or as other sound classes. Wheeze segmentation is a type of sound event detection or segmentation, i.e., the task of recognizing sound events and their respective temporal onsets and endings in a recording [10].

Our proposed method is grounded on the theoretical properties of harmonic-percussive source separation, which breaks down a signal into its harmonic and percussive components, and empirical mode decomposition (EMD), which decomposes a signal into intrinsic mode functions (IMFs). The main contributions of this paper encompass the method for wheeze segmentation as well as a comparison between our method and other state-of-the-art algorithms on a benchmark dataset, the ICBHI 2017 Respiratory Sound Database [11], [12]. Furthermore, a stratified analysis of the results identifies limitations of current methods and points out possible directions for future work.

The paper is structured as follows: in Section II, we provide an overview of state-of-the-art algorithms that have been used in similar works; in Section III, we present the database, as well as the details of the proposed method; in Section IV, the obtained results are analyzed; and, lastly, we conclude in Section V.

II. Related Work

Multiple systems for the automatic detection or segmentation of wheezes have been proposed in the literature, often reporting excellent results [13], [14], [15], [16], [17], [18], [19], [20], [21], [22]. A summary of state-of-the-art wheeze segmentation methods can be found in Table I. However, most works used small or private datasets containing a small number of wheezes and few sources of environmental noise, since a standard evaluation procedure has not been established [23]. In this paper, we evaluated our method on the ICBHI 2017 Respiratory Sound Database [11], [12], a large public database that has become a benchmark for the evaluation of algorithms analyzing respiratory sounds. Furthermore, we compared the performance of our method to three works: i) time-frequency wheeze detection (TFWD), the most cited work in the literature in this topic [13]; ii) wheeze signature in the spectrogram space (WSSS), a previous algorithm from our lab [18]; and iii) recursive approach via non-negative matrix factorization and Gini index sparsity (NMFG), a recent method [21]. Further details about these algorithms can be found below.

The first steps of the TFWD technique consist of sampling each recording at 5512 Hz and bandpass filtering the signal between 60 and 2100 Hz. Then, the spectrogram is computed using the short-time Fourier transform (STFT) with a 512 samples Hann window and 90% overlap. Subsequently, the underlying breath sound is subtracted from the total sound using a smoothing procedure based on box filtering. Then, peaks that exceed a specific magnitude threshold are selected, restricting the search to the interval of frequencies between 100 and 1000 Hz. Those peaks are then classified as wheezes or non-wheezes according to a set of criteria that include local maxima, peak coexistence, and continuity in time.

TABLE I
Summary of State-of-the-Art Wheeze Segmentation Methods

Ref.	Data	Methodology	Best Results
[13]	Participants: 13; Recordings: 13; Source: Private	Background subtraction; expert rules	Recall: 96%; Specificity: 94%
[15]	Participants: 28; Recordings: 28; Source: Public	Laplacian mask; multi-layer perceptron	Recall: 86%; Specificity: 83%
[18]	Participants: 12; Recordings: 24; Source: Private	Background subtraction; regularized weights	Recall: 91%; Specificity: 99%
[19]	Participants: 30; Recordings: 870; Source: Private	Ensemble EMD; instantaneous frequency; support vector machines	Precision; 95; Recall: 94%; Specificity: 94%
[20]	Participants: 16; Recordings: 16; Source: Private	Compressive sensing; hidden Markov models	Recall: 89%; Specificity: 96%
[21]	Participants: 32; Recordings: 32; Source: Private	Non-negative matrix factorization; clustering; spectral sparsity	Recall: 94%; Specificity: 97%

The WSSS method starts by filtering the signal with a Gaussian kernel. Subsequently, the spectrogram is computed using the STFT with a 128 ms flat top window and 75% overlap. Then, the same smoothing procedure used in TFWD to subtract the background is employed. Peaks above a certain threshold are then selected, restricting the search to the interval of frequencies between 100 and 1000 Hz. Afterwards, a geodesic morphological opening is applied to reduce the number of false positives. Finally, a binary array of weights with Gaussian regularization is computed, producing the final wheeze segments.

The NMFG approach comprises four stages. First, orthogonal non-negative matrix factorization bases (ONMF) are obtained from the normalized magnitude spectrogram. Then, the ONMF bases are clustered into two sets: bases that show higher periodicity, with energy concentrated in narrow-band spectral peaks, and bases that show lower periodicity, with energy distributed along the spectrum. In the third stage, the estimated wheezing spectrogram is improved by recursively factorizing new sets of ONMF bases to be re-clustered into wheeze bases or normal breath bases. Finally, the sparse behavior of the spectral energy distribution is analyzed to decide whether a sound excerpt contains wheezes.

III. Materials and Methods

A. Dataset

The ICBHI 2017 Respiratory Sound Database is a publicly available database with 920 audio files containing a total of 5.5 h of recordings acquired from 126 participants (79 males, 46 females, 1 unknown) of all ages (76 adults, 49 children, 1

TABLE II
DISTRIBUTION OF WHEEZES PER EQUIPMENT, AGE (RANGE, MEAN±STANDARD DEVIATION), SEX, BODY-MASS INDEX (RANGE), AND DIAGNOSIS

Stratification	Elements	Files	Wheeze Files	Wheezes
Equipment	AKG C417L	646	235	1231
	WelchAllyn Meditron	128	38	255
	3M Littmann 3200	60	31	190
	3M Littmann Classic II SE	86	37	222
Age (years)	Adults (19-93, 67.6±11.6)	838	321	1786
	Children (0-18, 4.9±4.6)	76	18	99
	Unknown	6	2	13
Sex	Female	323	99	475
	Male	591	240	1410
	Unknown	6	2	13
BMI	Normal weight (below 25)	326	123	682
	Overweight (25-29.9)	360	160	897
	Obese (above 30)	149	37	200
	Unknown	85	21	119
Diagnosis	Non-chronic (14 URTI, 2 LRTI, 6 bronchiolitis, 6 pneumonia)	75	23	121
	Chronic (64 COPD, 7 bronchiectasis, 1 asthma)	810	315	1774
	Healthy	35	3	3

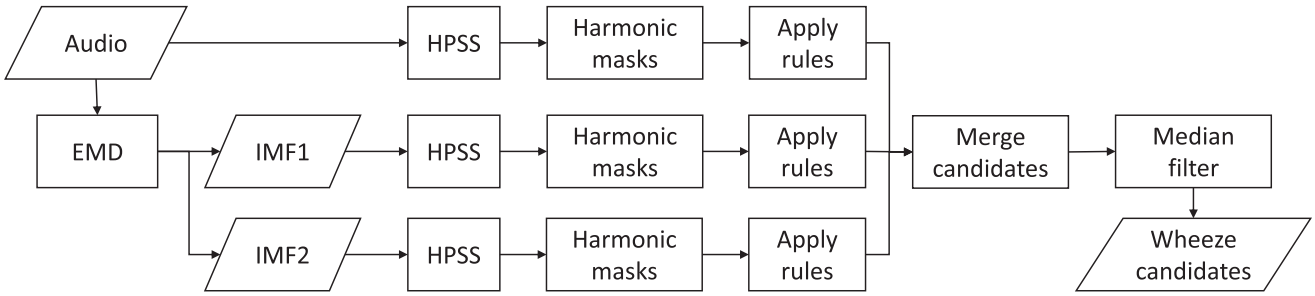

Fig. 1. Flowchart of the proposed method. EMD: empirical mode decomposition; HPSS: harmonic-percussive source separation; IMF: intrinsic mode frequency.

unknown) [11], [12]. The database contains audio samples collected independently by two research teams, over several years, in Portugal and Greece. The recordings were collected in clinical and home settings, using stethoscopes (WelchAllyn Meditron, 3 M Littmann 3200, 3 M Littmann Classic II SE) or microphones (AKG C417 L) with different sampling frequencies, and include various noise sources and background sounds. The database contains 1898 annotated wheezes, distributed among 341 audio files. The distribution of the number of files, number of files with annotated wheezes, and number of annotated wheezes per recording equipment, age, sex, body-mass index (BMI), and diagnosis is shown in Table II. To categorize by age, all participants under 18 were considered children. BMI categories were defined according to the World Health Organization guidelines [24]; as only three participants were underweight, a category merging normal and underweight was formed. Diagnosis classes were created by defining the chronic obstructive pulmonary disease (COPD), asthma, and bronchiectasis patients as chronic, and patients with lower respiratory tract infection (LRTI), upper respiratory tract infection (URTI), bronchiolitis, or pneumonia as non-chronic.

B. Proposed Method

Henceforth we present a method for wheeze segmentation. The method attaches harmonic-percussive separation to empirical mode decomposition with the goal of reducing the energy of other signals present in the respiratory sounds, such as crackles or handling noises, and increasing the salience of wheezes. This process is complemented by filters that adapt this framework to the specific properties of wheeze signals. A hyperparameter tuning, detailed in Section IV-B was performed to determine all the parameter values mentioned below. Fig. 1 shows a flowchart of the proposed method.

1) Preprocessing: The ICBHI Respiratory Sound Database contains recordings with different sampling rates. Therefore, we resampled every recording at 4000 Hz, the lowest sampling rate in the dataset.

As the database contains sounds with different duration and the segmentation method needed excerpts with fixed duration, we chopped the sounds into chunks of 10 s with an overlap of 90%.

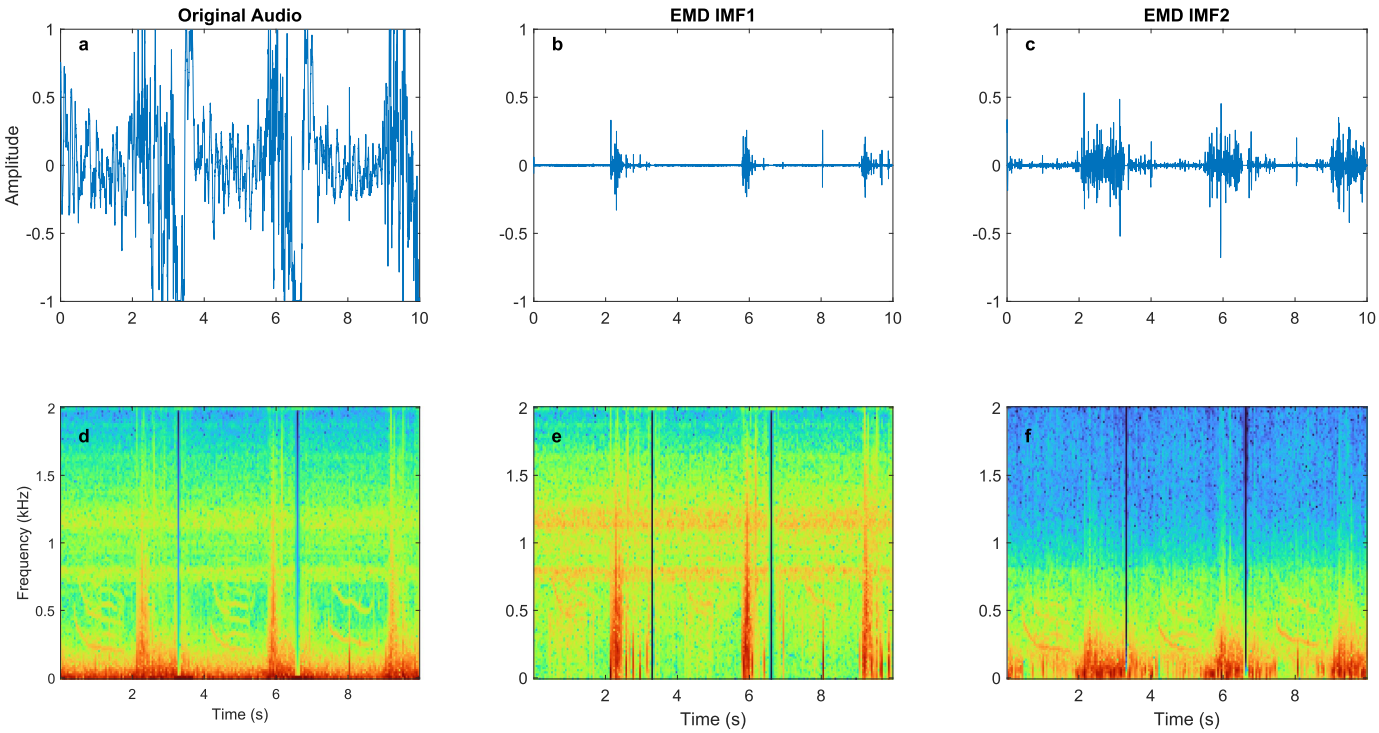

Fig. 2. (a), (d) Original audio waveform and spectrogram, respectively. (b), (e) EMD IMF1 waveform and spectrogram, respectively. (c), (f) EMD IMF2 waveform and spectrogram, respectively.

Then, we decomposed the signal into intrinsic mode functions (IMFs) using empirical mode decomposition (EMD) [25]. 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 [26]. We extracted the first two IMFs, as most of the wheezes' energy was concentrated on those IMFs. The process detailed below was carried out on three tracks: the original audio and the two IMFs. The result of the decomposition is shown in Fig. 2, as well as the respective spectrograms.

2) Harmonic Wheeze Segmenter: This segmentation method is based on harmonic-percussive source separation using median filtering [27]. Although the task of decomposing an audio signal into its harmonic and its percussive components has been used in many musical applications such as remixing or tempo estimation [28], it has not been applied in the segmentation of adventitious respiratory sounds. The idea that, in a magnitude spectrogram, broadband impulsive noises form stable vertical ridges and harmonics from pitched instruments form stable horizontal ridges is a useful approximation [29]. Since wheezes are continuous musical sounds with a dominant frequency and harmonics, they should as well produce stable horizontal ridges in the spectrogram. However, Driedger et al. [28] observed that this approach does not produce tight decompositions because some sounds are neither of clearly harmonic or percussive nature. Also, the leakage of harmonic sounds into the percussive component - and vice versa - depend on the parameter settings. Thus, they introduced an additional parameter, the separation factor, to tighten the harmonic-percussive separation. As our goal in this work was to detect and segment wheezes, we only needed to isolate the harmonic component. Fig. 3 depicts the various harmonic masks obtained for a particular excerpt of a recording.

First, we computed the spectrogram using the STFT with a 512 ms window and 87.5% overlap. Then, using median filtering in the horizontal (100 ms) and vertical (200 Hz) directions, we fetched harmonic- and percussive-enhanced spectrograms. Next, we obtained a harmonic binary hard mask by finding indices where the energy of the harmonic spectrum was at least 3 times the energy of the percussive spectrum, i.e., we used a separation factor of 3.

Then, we applied another median filter in the horizontal direction (100 ms) to clean the signal and, following Taplidou and Hadjileontiadis [13], grouped connected components to obtain the initial masks, as exhibited in Fig. 3(a), (b), (c). Subsequently, we eliminated components above 800 Hz (1600 Hz for tracheal sounds, an extra octave) and those that did not conform to the following rules to obtain the filtered masks, as displayed in Fig. 3(d), (e), (f):

- minimum duration (graphical width): 50 ms (typically, wheezes last more than 100 ms [3], but considering that the algorithm might not detect a full wheeze, half of that duration was deemed as the minimum duration);
- maximum duration (graphical width): 4 s (empirically set value, as the literature mentions no maximum wheeze duration);
- minimum frequency range (graphical height): 10 Hz (to avoid components with no spectral spread);

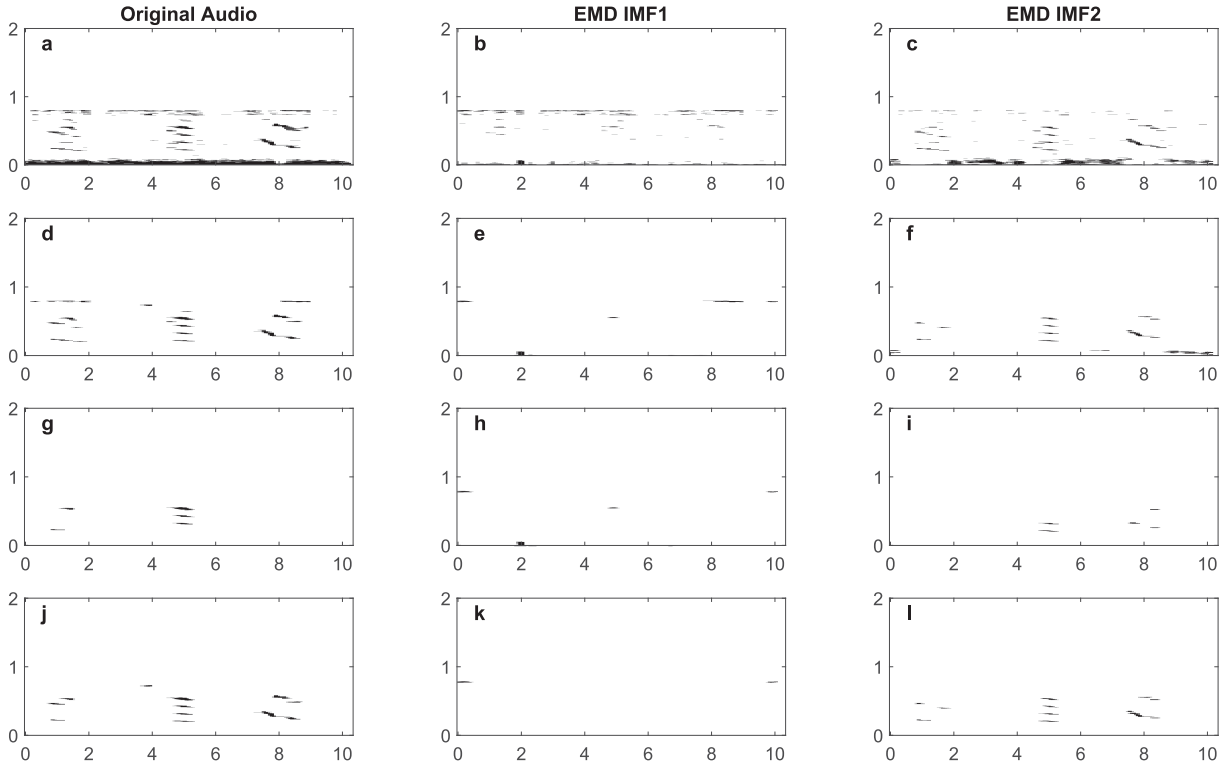

Fig. 3. (a), (d), (g), (j) Initial mask, filtered mask, top mask, and final mask, respectively, for original audio track. (b), (e), (h), (k): initial mask, filtered mask, top mask, and final mask, respectively, for EMD IMF1 track; (c), (f), (i), (l): initial mask, filtered mask, top mask, and final mask, respectively, for EMD IMF2 track.

- maximum frequency range (graphical height): 300 Hz (to avoid components with too much spectral spread);
- maximum graphical perimeter: graphical area, i.e., perimeter should be smaller than area to guarantee that the selected components were longilineal;
- maximum graphical extent: 1, i.e., component area should be smaller than bounding box area;
- minimum graphical orientation: 5° (to avoid strictly horizontal components).

Next, we computed the ratio between perimeter and area and sorted the remaining components from smallest to largest ratio between perimeter and area. At this point, at most 5 components (1 per each 2 s) were selected to obtain the top masks, as shown in Fig. 3(g), (h), (i). This ratio and the selection of the largest components were chosen because components with longilineal forms are less likely to be spurious components derived of a poor harmonic-percussive separation, following the observation that longer contours are less likely to be errors of a melody extraction algorithm [30]. The top mask components were used to compute the mean and the standard deviation of the logarithm with base 2 of the graphical area, centroid, width, and orientation. Finally, those components with properties within mean ±1.5 standard deviations and centroid above 100 Hz (typically, the minimum wheeze frequency) were selected. The goal of this step was to select components with common characteristics, as wheezes in a given excerpt should be homogeneous. The final mask was obtained from the selected components, as illustrated in Fig. 3(j), (k), (l).

After retrieving the final mask, we merged components that were present during the same time frames. The final output comprised the beginnings and endings of each wheeze candidate.

3) Postprocessing: In this step, we merged the wheeze candidates of the three outputs, keeping only the ones that appeared in at least two tracks. A median filter with the length of the median wheeze duration (400 ms) was applied to delete spurious candidates. Fig. 4 shows details about the postprocessing.

IV. Evaluation

A. Concepts and Measures

Before presenting the results, some relevant concepts are defined below:

- *Annotated Event (AE):* time boundaries that mark the beginning and ending of a wheeze, as decided by the annotator
- *Segmented Event (SE):* time boundaries that mark the beginning and ending of a wheeze candidate, as decided by the segmentation algorithm
- *Detected Event (DE):* AE that is detected by the algorithm
- *Undetected Event (UE):* AE that is not detected by the algorithm
- *False Event (FE):* SE whose beginning and ending are outside the boundaries of an AE

Likewise, a measure of similarity and a threshold level are needed to define what constitutes a detected event (DE) or undetected event (UE). We used two common measures of

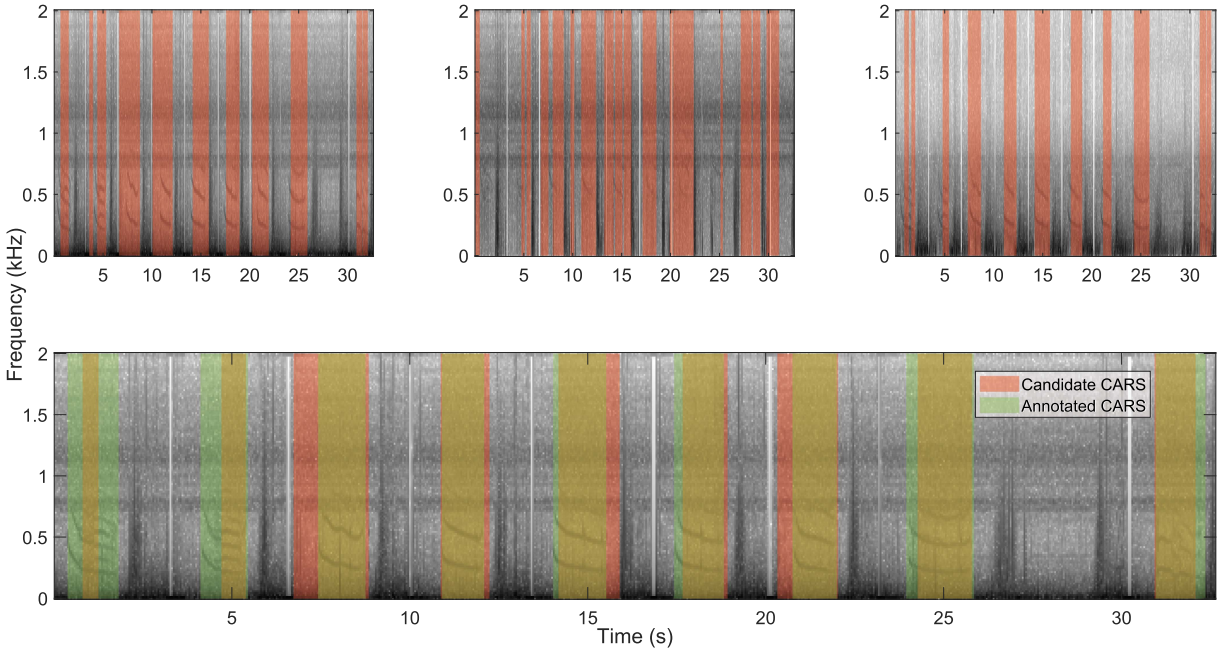

Fig. 4. Top Left: wheeze candidates (red) for original audio track; Top Center: wheeze candidates (red) for EMD IMF1 track; Top Right: wheeze candidates (red) for EMD IMF2 track; Bottom: merged wheeze candidates (red) after applying median filter with a length of 400 ms, and annotated wheeze (green).

similarity, the Jaccard Index (JI) and the Overlap Coefficient (OC),

$$JI(A,B) = \frac{A \cap B}{A \cup B} \tag{1}$$

$$OC(A,B) = \frac{A \cap B}{\min(|A|,|B|)} \tag{2}$$

and two threshold levels: a loose threshold (10%), i.e., at least 10% of an annotated event coincided with a segmented event, and a strict threshold (50%), i.e., at least 50% of an annotated event coincided with a segmented event. Given these concepts, we can define relevant evaluation measures:

$$Precision = \frac{DE}{(DE + UE)} \tag{3}$$

$$Recall = \frac{DE}{(DE + FE)} \tag{4}$$

$$F1 = \frac{(2 \times Precision \times Recall)}{(Precision + Recall)} \tag{5}$$

In this section, we reported only F1 values, but tables containing Precision and Recall values can be found in the Supplementary Material.

B. Hyperparameter Tuning

We performed hyperparameter tuning on a small dataset with 15 files from 15 subjects to find the parameter values mentioned in Section III-B. This dataset includes sounds from 15 participants, 13 adults (6 with COPD, 3 with pneumonia, 2 with LRTIs, 1 with asthma, and 1 with bronchiectasis) and 2 children (1 healthy and 1 with bronchiolitis). The sounds were acquired using the WelchAllyn Meditron Stethoscope (6 participants), the Littmann Classic II SE Stethoscope (6 participants), and the AKG C417 L Microphones (3 participants). This dataset contains 85 wheezes distributed among 10 files and 5 files with no wheezes. Two goals were set: maximum F1 given an OC threshold of 0.5, i.e., at least 50% of an annotated event coincided with a segmented event using the overlap similarity; maximum F1 given a JI threshold of 0.5, i.e., at least 50% of an annotated event coincided with a segmented event using the Jaccard similarity. Henceforth, the resulting algorithms are referred to as HWS_OC_50, and HWS_JI_50, respectively.

TABLE III
F1 FOR DIFFERENT THRESHOLD VALUES IN THE FULL DATASET

Algorithm	10% OC	10% JI	50% OC	50% JI
TFWD	0.162	0.139	0.154	0.071
WSSS	0.368	0.354	0.347	0.179
NMFG	0.270	0.149	0.234	0.024
HWS_JI_50	0.397	0.378	0.363	0.176
HWS_OC_50	**0.419**	**0.401**	**0.384**	**0.182**

C. Segmentation Baselines

The results presented in this section compare our algorithms with the aforementioned three baseline methods: time-frequency wheeze detection (TFWD) [13]; wheeze signature in the spectrogram space (WSSS) [18]; recursive approach via non-negative matrix factorization and Gini index sparsity (NMFG) [21].

D. Overall Results

Table III presents the overall F1 of the aforementioned algorithms for two threshold values (10% and 50%) and two

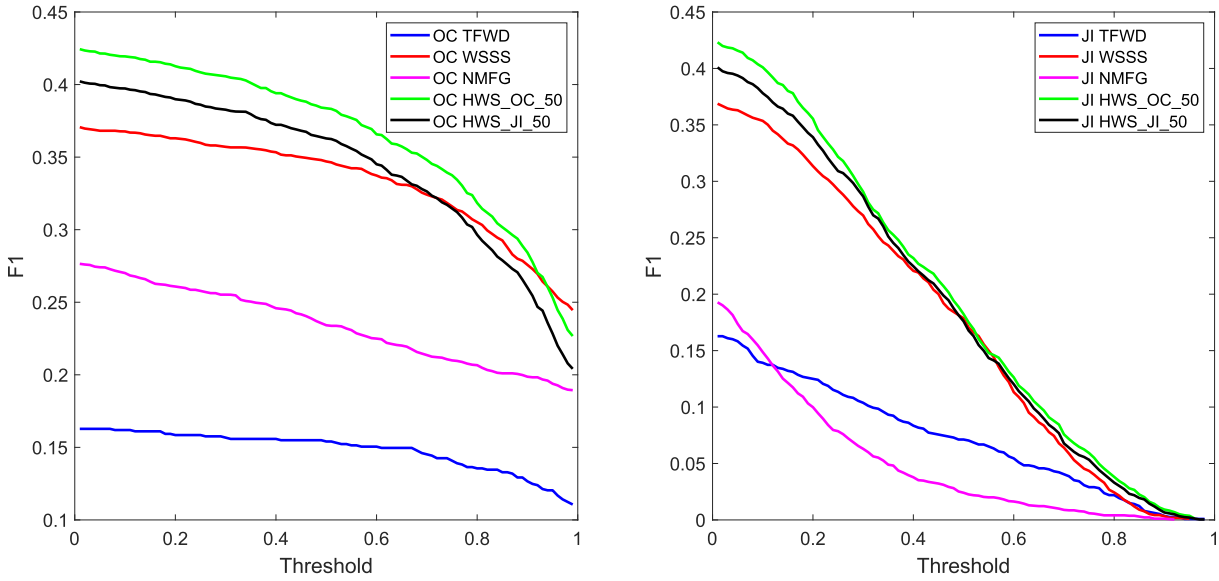

Fig. 5. Threshold-dependent results with overlap similarity (left) and Jaccard similarity (right).

TABLE IV
F1 FOR DIFFERENT THRESHOLD VALUES IN WHEEZE FILES

Algorithm	10% OC	10% JI	50% OC	50% JI
TFWD	0.168	0.144	0.160	0.074
WSSS	0.480	0.461	0.453	0.233
NMFG	0.391	0.216	0.338	0.035
HWS_JI_50	0.532	0.504	0.485	0.234
HWS_OC_50	**0.545**	**0.521**	**0.499**	**0.237**

similarity measures (OC and JI). The best F1, 41.9%, was reached by our proposed algorithm HWS_OC_50 with the 10% OC threshold. The results were slightly lower with the 10% JI threshold, with the best algorithm, HWS_OC_50, attaining an F1 of 40.1%. Increasing the threshold to 50% using the JI, the results decreased substantially, with HWS_OC_50 achieving the best F1, 18.2%. Using the 50% OC threshold, the results did not decrease as much as with the JI. The best F1, 38.4%, was reached by HWS_OC_50, the algorithm that achieved the best results for all the thresholds. The disparity between these results and those reported in the literature is striking. We believe that earlier datasets were too small and lacked environmental noise, leading to poor generalizability of the algorithms developed using those datasets. In fact, as mentioned above, the recordings of the ICBHI Respiratory Sound Database were collected in clinical and home settings, i.e., in realistic scenarios, a significant challenge to the segmentation algorithms, as shown by the attained results.

E. Results in Wheeze Files

A plausible scenario for wheeze segmentation would be to consider only files containing annotated wheezes, as an automatic method could be used to segment wheezes in sounds from patients with respiratory conditions where wheezes are the main adventitious respiratory sounds, e.g., asthma. In this scenario, the precision of all algorithms would increase while the recall would remain unchanged. Table IV shows the F1 values for the aforementioned algorithms when considering only files containing annotated wheezes. As in Section IV-D, there is a significant drop in performance when using JI with a threshold of 50%. HWS_OC_50 attained the best F1 with all the thresholds, reaching 54.5% with the 10% OC threshold, 52.1% with the 10% JI threshold, 49.9% with the 50% OC threshold, and 23.7% with the 50% JI threshold.

F. Threshold-Dependent Results

Additionally, we analyzed how the chosen similarity measures (OC and JI) and different threshold levels affected the results. Fig. 5 shows the results for OC and JI with the threshold varying between 0.01 and 0.99 in 0.01 steps. As expected, the JI curves are much steeper than the OC curves. Still, HWS_OC_50 was consistently equal to or better than the alternatives except when surpassed by WSSS for OC above 94%. This could have happened because the postprocessing median filter, while important to lower the number of false positives, may reduce the correct number of samples at the extremities of a wheeze candidate. Although JI is widely used for segmentation tasks, we believe OC is more relevant for this problem, as JI is insensitive to the length of the segments [31]. Additionally, depending on the goal defined by the clinician, small thresholds can be clinically appropriate, e.g., if the precise boundaries of a wheeze are not clinically relevant.

G. Stratified Results

In addition to the previous analyses, it is important to investigate how our method performed when considering patients from different demographics. Therefore, we stratified the participants according to recording equipment, age, sex, BMI, and diagnosis, and computed the F1 scores of the TFWD, WSSS, NMFG, HWS_JI_50, and HWS_OC_50 algorithms for each of those variables, as shown in Table V. The table only considers the 10%

TABLE V
10% OC F1 PER EQUIPMENT, AGE, SEX, BODY-MASS INDEX, AND DIAGNOSIS

Stratification	Elements	TFWD	WSSS	NMFG	HWS_JI_50	HWS_OC_50
Equipment	AKG C417L	0.151	0.336	0.253	0.376	**0.385**
	WelchAllyn Meditron	0.140	**0.469**	0.253	0.343	0.377
	Littmann Classic II SE	0.209	0.571	0.369	0.757	**0.777**
	Littmann 3200	0.203	0.182	0.363	0.333	**0.375**
Age	Children	0.019	0.429	0.110	0.507	**0.522**
	Adults	0.170	**0.480**	0.397	0.393	0.415
Sex	Female	0.303	0.343	0.270	0.430	**0.450**
	Male	0.110	0.378	0.272	0.388	**0.410**
BMI	Normal Weight	0.186	0.315	0.296	0.422	**0.433**
	Overweight	0.154	0.425	0.308	0.443	**0.472**
	Obese	0.183	**0.287**	0.147	0.196	0.211
Diagnosis	Non-chronic	0.164	0.340	0.176	0.571	**0.590**
	Chronic	0.162	0.371	0.282	0.391	**0.413**

OC threshold. Other results can be found in the Supplementary Material.

Regarding equipment, WSSS achieved the best F1 for Meditron, 46.9%, and HWS_OC_50 reached the best F1 for the other equipment categories, 38.5% for AKG C417 L, 37.7% for Littmann 3200, and 77.7% for Littmann Classic II SE. All the algorithms reached their best results with the Littmann Classic II SE. Presence of other sounds, e.g. cough, annotator consistency, or noise levels could be relevant factors to justify this disparity in the performance and should be investigated in future work, especially considering the comparatively very high results attained by the HWS methods with the Littmann Classic II SE.

Regarding age, WSSS and HWS_OC_50 reached F1 scores above 40% for both children and adults, with WSSS achieving the best F1 for adults, 48%, and HWS_OC_50 achieving the best F1 for children, 52.2%. The performance of TFWD was especially poor in children's recordings, reaching an F1 of 1.9%. This might be due to the dataset for which it was optimized not containing any children's respiratory sounds or any sounds recorded with the Meditron, which was the equipment used to acquire all the children's sounds in the ICBHI 2017 Respiratory Sound Database.

HWS_OC_50 attained the best F1 for both sex categories, reaching 45% in female participants and 41% in male participants. The database is skewed towards male participants, with female participants accounting for 35% of the recordings, 29% of the files containing annotated wheezes, and 25% of the wheezes. Furthermore, sounds from female participants are overrepresented in all the weight categories except obese, and respiratory sounds in obese people typically have lower intensity and can be inaudible during quiet breathing [32].

In what concerns the results according to diagnosis, the healthy category is not shown; no algorithm detected any of the 3 wheezes, therefore precision and recall were 0. HWS_OC_50 achieved the best F1 score for both acute and chronic categories. Nevertheless, while WSSS and TFWD achieved similar F1 scores for both categories, there was a substantial disparity in the other algorithms, with HWS_OC_50 reaching F1 scores of 59% and 41.3% for acute and chronic wheezes, respectively. We speculate that the equipment and the age might be the main factors for this disparity, as all the sounds from children with non-chronic conditions were acquired with the Meditron and all the sounds from adults with acute conditions were acquired with the Littmann Classic II SE.

H. Results in Clean Databases

In addition to the evaluation on the ICBHI 2017 Respiratory Sound Database, we conducted supplementary analyses on two *clean* datasets. First, we annotated the precise locations of wheezes in 15 noise-free files of the Jordanian respiratory sound database [33]. Then, we annotated 4 public files of the R.A.L.E. database [34] containing wheezes. In both cases, the NMFG algorithm achieved the best results, reaching F1 scores of 90.3% and 98.1% in the Jordanian and the R.A.L.E. clean databases, respectively. Detailed results can be found in the supplementary materials.

I. Computational Complexity

Regarding computational complexity, we ran each algorithm on the Jordanian clean dataset and estimated their computation time. The times were the following: TFWD: 2 s; WSSS: 1 s; NMFG: 17 s; HWS_JI_50: 140 s; HWS_OC_50: 175 s. All methods were implemented using MATLAB. The NMFG algorithm was run on a Intel(R) Core(TM) i9-12900HK CPU @ 2.90 GHz with 64 GB of RAM, while the rest of the algorithms were run on a Intel(R) Core(TM) i7-8750H CPU @ 2.20 GHz with 32 GB of RAM. Given their recursive nature, the HWS algorithms needed much more time to run than the other algorithms. However, parallel computing could considerably speed up the process.

V. CONCLUSION

Contrary to what has been reported in the literature, we found that the problem of automatic wheeze segmentation has not been solved for real life scenario applications, i.e., when environmental noise and other confounding sounds affect the quality of

respiratory sounds. This work demonstrates the importance of benchmark datasets for the evaluation of new algorithms.

Nonetheless, it is important to note the shortcomings of the ICBHI 2017 Respiratory Sound Database, namely the lack of gold standard annotations, i.e., wheezes annotated by various health professionals, or the absence of sounds from healthy adults.

Yet, our analysis has shown that adapting existing systems to particular demographic characteristics of patients might be a promising route - body size, for example, affects the signal in ways that can be addressed at the data acquisition stage, and the choice of recording equipment dictates subsequent signal processing decisions. That adaptation could be a stepping-stone in the direction of data acquisition and algorithm personalization, which we believe would make these automatic methods clinically viable. Another possible path for future work would be to increase the existing databases and to try deep learning architectures, leveraging their power to learn complex patterns from large datasets. Architectures that capture temporal dependencies could be especially relevant, even though their lack of interpretability might be a limitation in the context of clinical applications.

Acknowledgment

We would like to thank Luís Mendes for sharing the WSSS code and Juan de la Torre Cruz for running the NMFG code. Furthermore, we are thankful to LASI - Laboratório Associado em Sistemas Inteligentes.

References

[1] W. W. Labaki and M. L. K. Han, "Chronic respiratory diseases: A global view," *Lancet Respir. Med.*, vol. 8, no. 6, pp. 531–533, 2020.

[2] WHO, "The top 10 causes of death," 2020. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

[3] A. Bohadana, G. Izbicki, and S. S. Kraman, "Fundamentals of lung auscultation," *New England J. Med.*, vol. 370, no. 8, pp. 744–751, Feb. 2014.

[4] A. Marques, A. Oliveira, and C. Jácome, "Computerized adventitious respiratory sounds as outcome measures for respiratory therapy: A systematic review," *Respir. Care*, vol. 59, no. 5, pp. 765–776, 2014.

[5] A. R. Watson, R. Wah, and R. Thamman, "The value of remote monitoring for the COVID-19 pandemic," *Telemed. e-Health*, vol. 26, no. 9, pp. 1110–1112, 2020.

[6] A. R. A. Sovijärvi, F. Dalmasso, J. Vanderschoot, L. P. Malmberg, G. Righini, and S. A. T. Stoneman, "Definition of terms for applications of respiratory sounds," *Eur. Respir. Rev.*, vol. 10, no. 77, pp. 597–610, 2000.

[7] L. J. Hadjileontiadis and Z. M. K. Moussavi, *Current Techniques for Breath Sound Analysis*. Cham, Switzerland: Springer, 2018, pp. 139–177.

[8] H. Pasterkamp, P. L. P. Brand, M. Everard, L. Garcia-Marcos, H. Melbye, and K. N. Priftis, "Towards the standardisation of lung sound nomenclature," *Eur. Respir. J.*, vol. 47, no. 3, pp. 724–732, 2016.

[9] A. Marques and A. Oliveira, *Normal Versus Adventitious Respiratory Sounds*. Cham, Switzerland: Springer, 2018, ch. 10, pp. 181–206.

[10] S. Adavanne and T. Virtanen, "A report on sound event detection with different binaural features," 2017, *arXiv:1710.02997*.

[11] B. M. Rocha et al., "A respiratory sound database for the development of automated classification," in *Proc. Int. Federation Med. Biol. Eng.*, 2018, vol. 66, pp. 33–37.

[12] B. M. Rocha et al., "An open access database for the evaluation of respiratory sound classification algorithms," *Physiol. Meas.*, vol. 40, no. 3, 2019, Art. no. 035001.

[13] S. A. Taplidou and L. J. Hadjileontiadis, "Wheeze detection based on time-frequency analysis of breath sounds," *Comput. Biol. Med.*, vol. 37, no. 8, pp. 1073–1083, 2007.

[14] J.-C. Chien, H.-D. Wu, F.-C. Chong, and C.-I. Li, "Wheeze detection using cepstral analysis in Gaussian mixture models," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol.*, 2007, pp. 3168–3171.

[15] R. J. Riella, P. Nohama, and J. M. Maia, "Method for automatic detection of wheezing in lung sounds," *Braz. J. Med. Biol. Res.*, vol. 42, no. 7, pp. 674–684, Jul. 2009.

[16] J. Zhang, W. Ser, J. Yu, and T. T. Zhang, "A novel wheeze detection method for wearable monitoring systems," in *Proc. Int. Symp. Intell. Ubiquitous Comput. Educ.*, 2009, pp. 331–334.

[17] M. Bahoura, "Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes," *Comput. Biol. Med.*, vol. 39, no. 9, pp. 824–843, 2009.

[18] L. Mendes et al., "Detection of wheezes using their signature in the spectrogram space and musical features," in *Proc.37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2015, pp. 5581–5584.

[19] M. Lozano, J. A. Fiz, and R. Jané, "Automatic differentiation of normal and continuous adventitious respiratory sounds using ensemble empirical mode decomposition and instantaneous frequency," *IEEE J. Biomed. Health Inform.*, vol. 20, no. 2, pp. 486–497, Mar. 2016.

[20] D. Oletic and V. Bilas, "Asthmatic wheeze detection from compressively sensed respiratory sound spectra," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 5, pp. 1406–1414, Sep. 2018.

[21] J. De La Torre Cruz, F. J. C. Quesada, J. J. C. Orti, P. V. Candeas, and N. R. Reyes, "Combining a recursive approach via non-negative matrix factorization and gini index sparsity to improve reliable detection of wheezing sounds," *Expert Syst. Appl.*, vol. 147, 2020, Art. no. 113212.

[22] A. Semmad and M. Bahoura, "Long short term memory based recurrent neural network for wheezing detection in pulmonary sounds," in *Proc. IEEE Int. Midwest Symp. Circuits Syst.*, 2021, pp. 412–415.

[23] R. X. A. Pramono, S. Bowyer, and E. Rodriguez-Villegas, "Automatic adventitious respiratory sound analysis: A systematic review," *PLoS ONE*, vol. 12, no. 5, 2017, Art. no. e0177926.

[24] WHO, "Body mass index - BMI," 2022. [Online]. Available: https://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi

[25] N. E. Huang et al., "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. A: Math., Phys. Eng. Sci.*, 1998, vol. 454, no. 1971, pp. 903–995.

[26] G. Rilling and P. Flandrin, "One or two frequencies? The empirical mode decomposition answers," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 85–95, Jan. 2008.

[27] D. Fitzgerald, "Harmonic/percussive separation using median filtering," in *Proc. 13th Int. Conf. Digit. Audio Effects*, 2010, pp. 1–4.

[28] J. Driedger, M. Müller, and S. Disch, "Extending harmonic-percussive separation of audio signals," in *Proc. 15th Int. Soc. Music Inf. Retrieval Conf.*, 2014, pp. 611–616.

[29] D. Fitzgerald and M. Gainza, "Single channel vocal separation using median filtering and factorisation techniques," *ISAST Trans. Electron. Signal Process.*, vol. 4, no. 1, pp. 62–73, 2010.

[30] J. Salamon, B. Rocha, and E. Gómez, "Musical genre classification using melody features extracted from polyphonic music signals," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2012, pp. 81–84.

[31] B. Rees, "Similarity in graphs: Jaccard versus the overlap coefficient," 2021. [Online]. Available: https://medium.com/rapids-ai/similarity-in-graphs-jaccard-versus-the-overlap-coefficient-610e083b877d

[32] N. Gavriely and D. Cugell, *Breath Sounds Methodology*. Boca Raton, FL, USA: CRC Press, 1995.

[33] M. Fraiwan, L. Fraiwan, B. Khassawneh, and A. Ibnian, "A dataset of lung sounds recorded from the chest wall using an electronic stethoscope," *Data Brief*, vol. 35, 2021, Art. no. 106913.

[34] D. Owens, "R.A.L.E. Lung sounds 3.0," *CIN: Comput., Inform., Nurs.*, vol. 5, no. 3, pp. 9–10, 2002.