Biomedical Signal Processing and Control 87 (2024) 105451

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Ensemble deep learning model for dimensionless respiratory airflow estimation using respiratory sound

Diogo Pessoa [a,*], Bruno Machado Rocha [a], Maria Gomes [b], Guilherme Rodrigues [b], Georgios Petmezas [c], Grigorios-Aris Cheimariotis [c], Nicos Maglaveras [c], Alda Marques [b,d], Inéz Frerichs [e], Paulo de Carvalho [a], Rui Pedro Paiva [a]

[a] University of Coimbra, Centre for Informatics and Systems of the University of Coimbra, Department of Informatics Engineering, Coimbra, 3030-290, Portugal
[b] Lab3R — Respiratory Research and Rehabilitation Laboratory, School of Health Sciences (ESSUA), University of Aveiro, Aveiro, 3810-193, Portugal
[c] 2nd Department of Obstetrics and Gynaecology, Laboratory of Computing, Medical Informatics and Biomedical-Imaging Technologies, Medical School, Aristotle, University of Thessaloniki, Thessaloniki, 54124 Greece
[d] Institute of Biomedicine (iBiMED), University of Aveiro, Aveiro, 3810-193, Portugal
[e] Department of Anaesthesiology and Intensive Care Medicine, University Medical Centre Schleswig-Holstein, Campus Kiel, Kiel, 24105, Germany

ARTICLE INFO

Dataset link: https://data.mendeley.com/datasets/f43c7snks5/1, https://github.com/DiogoMPessoa/Dimensionless-Respiratory-Airflow-Estimation

Keywords:
Respiratory sound analysis
Electrical impedance tomography
Dimensionless respiratory airflow
Flow–sound relationship
Acoustical airflow estimation
Deep learning

ABSTRACT

In recent years, computerized methods for analyzing respiratory function have gained increased attention within the scientific community. This study proposes a deep-learning model to estimate the dimensionless respiratory airflow using only respiratory sound without prior calibration. We developed hybrid deep learning models (CNN + LSTM) to extract features from the respiratory sound and model their temporal dependencies. Then, we used an ensemble approach to combine multiple outputs of our models and obtain the respiratory airflow waveform for entire respiratory audio signals as the final output. We conducted a comprehensive set of experiments and evaluated the models using several regression evaluation metrics to assess how the models would perform in various circumstances of different complexity. The methods were developed and evaluated considering respiratory sound and electrical impedance tomography (EIT) data from 50 respiratory patients (15 female and 35 male with an average age of 67.4 ± 8.9 years and body mass index of 27.8 ± 5.6 kg/m^2). An external assessment was conducted using an external database, the Respiratory Sound Database (RSD). This was an indirect evaluation because the RSD does not provide the ground truth values of the dimensionless respiratory airflow. In the most complex evaluation task (*Task II*), we achieved the following results for the estimation of the normalized dimensionless respiratory airflow curve: mean absolute error = 0.134 ± 0.061; root mean squared error = 0.170 ± 0.075; dynamic time warping similarity = 3.282 ± 1.514; Pearson correlation coefficient = 0.770 ± 0.235. External assessment with the RSD showed that the performance of our model decreased when devices different from the ones used for their training were considered. Our study demonstrated that deep learning models could reliably estimate the dimensionless respiratory airflow.

1. Introduction

Respiratory diseases are among the most significant causes of morbidity and mortality worldwide and are responsible for a substantial strain on individuals, healthcare systems, and society [1,2]. Early diagnosis and frequent monitoring are essential for the management of these patients. Currently, chronic respiratory diseases are not curable; however, various pharmacological (e.g., bronchodilators) and non-pharmacological (e.g., physical activity, pulmonary rehabilitation) treatments contribute to the improvement of the symptoms (e.g., shortness of breath, fatigue), physical and emotional function, and quality of life of people with such diseases. Nevertheless, early diagnosis, detection of acute exacerbation (defined as an acute worsening of respiratory symptoms that result in additional therapy [3]), and long-term management remain highly challenging and have led to significant research efforts to improve the prognosis of these conditions. One of the most active research areas has been the computerized analysis of respiratory sounds. The main objective of these techniques is to overcome some of the drawbacks of conventional methods and provide more objective methods to monitor and diagnose patients suffering from lung diseases [4].

Lung auscultation is one of the most commonly used techniques by clinicians when performing routine physical examinations [5,6].

* Corresponding author.
E-mail address: dpessoa@dei.uc.pt (D. Pessoa).

https://doi.org/10.1016/j.bspc.2023.105451
Received 8 March 2023; Received in revised form 31 July 2023; Accepted 12 September 2023
Available online 5 October 2023

Although other measures are available to diagnose and monitor respiratory diseases (e.g., spirometry and medical imaging techniques), the information derived from respiratory sounds differs and complements these measures [7]. When performing lung auscultation, clinicians usually look up for the presence of adventitious respiratory sounds. These are additional respiratory sounds superimposed on normal respiratory sounds, and their presence is usually suggestive of a respiratory disorder [8,9]. Their phase (inspiratory/expiratory) and location within each phase (early/mid/late inspiratory or expiratory) of the respiratory cycle are other important parameters of clinical interest to identify the respiratory status of a patient and allow the differential diagnosis of various cardiorespiratory pathologies [5,10]. For instance, early inspiratory crackles at the lung bases have been found to be an indicator strongly correlated with chronic obstructive pulmonary disease (COPD) [11].

Characteristics of respiratory sounds change with gender, location of auscultation site, body size, posture, and airflow rate. In respiratory sound research, the simultaneous measurement of airflow and respiratory sounds is essential [12]. Respiratory airflow is defined as the volumetric flow rate of air inhaled and exhaled as a function of time [13]. The airflow signal distinguishes between respiratory phases (inspiration and expiration) with exact absolute values of volumetric changes during these phases, and it provides information about the current state of the lung [12]. In our study, we proposed a novel calculation of dimensionless respiratory airflow curves from respiratory sound recordings. As the name suggests, the dimensionless respiratory airflow is a measure similar to regular airflow; however, it does no provides any information regarding the absolute volume of air being inhaled or exhaled at any given time. Therefore, the dimensionless airflow curve allows the assessment of a subject inspiration and expiration patterns with relative flow rates.

The relationship between respiratory sounds and respiratory flow can also reveal the pathophysiology of the respiratory system and can be used as a basis for acoustical airflow estimation [13]. Currently, the state-of-the-art techniques used for measuring airflow are spirometry and pneumotachography [14]. Even though the methods can accurately estimate the respiratory flow, they require complex setups, namely a mouthpiece, and are unsuitable for continuous monitoring, particularly in telemonitoring applications.

With recent technological advances, the research and development of computerized methods for the automatic analysis of respiratory sounds have intensified with the availability of electronic stethoscopes (or similar acquisition devices). One of the most promising applications is the use of an electronic stethoscope paired with an application for respiratory sound processing and analysis [15]. Such setups enable the deployment of algorithms to estimate respiratory airflow in actual clinical settings and telemonitoring applications. Moreover, it may allow for continuous monitoring, overcoming some drawbacks of spirometry and pneumotachography.

Our main objective with this work was to develop a method to estimate the dimensionless respiratory airflow using only respiratory sound without prior knowledge or calibration. To this end, we proposed a hybrid deep neural network to model this airflow curve. The proposed models were developed from patient data obtained in a prospective clinical study. Moreover, we carried out a comprehensive set of experiments, even utilizing an external large respiratory sound database, to validate their performance and suitability for the intended purpose. It should also be noted that in this work, we used electrical impedance tomography (EIT) to obtain the dimensionless respiratory airflow (see Section 3.2).

EIT is a non-invasive, radiation-free imaging technique that relies on the application of alternating electrical currents on the external surface of the body to assess its internal electrical characteristics and generate bio-impedance/conductivity images/maps [16,17]. Previous clinical studies have supported the validity and reproducibility of EIT findings by comparing them against reference techniques such as CT-scan, single-photon emission CT, positron emission tomography, vibration response imaging, inert-gas washout, and spirometry [17]. Moreover, EIT is an approved medical method [18,19]. The global EIT waveform (dimensionless respiratory airflow curve) has also been used to determine breathing patterns [20].

The main contributions of this work can be highlighted in the following key points:

- Proposal of a new airflow concept based on the global electrical impedance tomography waveform (dimensionless respiratory airflow);
- First deep-learning-based method for airflow estimation;
- Method developed and evaluated on a large number of subjects with multiple scenarios of different complexity (with internal and external evaluation).

The article is organized as follows: (1) Introduction: presentation of the article's context, main motivation, and relevance; (2) Related Work: presentation of several related works published in the area as well as some considerations; (3) Materials and Methods: presentation of the data as well as the methodology used in this work; (4) Results: presentation of the obtained results; (5) Discussion: discussion of several aspects of the work, namely the results, strengths, and limitations; (6) Conclusion: final remarks and possible directions for future work.

2. Related work

The respiratory acoustic analysis allows the assessment of changes in respiratory sounds, aiding the diagnosis and management of respiratory diseases [13]. Over the years, several studies have attempted to explain the correlation between respiratory sounds and the airflow generated in the airway system, mainly based on classical signal processing analysis [13]. In fact, one of the most explored topics in the area of flow–sound relationship is the breathing phase-detection solely using respiratory sound, and the potential use of deep learning methods has been proposed to tackle this issue [5,21].

Although previous works had already shown a strong correlation between respiratory sounds and airflow [13], it was not until 2002 that it was attempted to derive the acoustical respiratory flow from respiratory sounds, with the respective error estimation [22]. Using an exponential model, the authors estimated the respiratory airflow from the average power of normal tracheal sounds collected from 10 healthy individuals at different flow rates. The error was evaluated by comparing the estimated airflow with the actual airflow measured using a pneumotachograph and found to be 5.8±3.0% when compared to the target airflow. However, that model required subject-specific calibration.

A novel method for estimating airflow using the tracheal sound entropy and the correlation between airflow and the entropy of respiratory sounds has been examined in [23]. Airflow and respiratory sounds were recorded from 10 healthy subjects, and three different models were studied to identify the best model for estimating airflow from the entropy of tracheal sounds. The authors found an overall estimation error of 8.3±2.8% and 9.6±2.8% for inspiration and expiration phase detection, respectively. However, their technique still required one breath cycle to calibrate the flow estimation model. Later, the same authors developed a new method that did not require previous calibration [24]. Tracheal sounds and airflow signals were simultaneously recorded from 93 healthy individuals, both smokers and non-smokers. This method was based on the relationship between flow and sound power. Results showed that flow estimation error based on the group-calibrated model was less than 10%, and the authors claimed that their model could estimate the respiratory flow in subjects with similar anthropometric features without needing to calibrate the model parameters for every individual.

More recently, a study confirmed that airflow could be estimated through acoustical means from respiratory sounds without previous knowledge of the respiratory phases and without needing an additional algorithm for phase detection [12]. The authors have used a 16-channel

recording device paired with a pneumotachograph to simultaneously record the lung sounds on the posterior chest and the respiratory flow. A total of 6 healthy subjects were recorded in a supine position. Then, the authors extracted the linear frequency cepstral coefficient (LFCC) features and mapped these on the airflow signal using multivariate polynomial regression to perform acoustic airflow estimation. Results suggested that the acoustical airflow during inhalation and accuracy of breath phase detection could be better estimated at higher airflow rates.

A non-invasive instrument capable of estimating respiratory flow parameters through tracheal sound analysis has also been developed [25]. That study indicated that the tracheal sound entropy closely followed the variation in airflow, i.e., the measured airflow was highly correlated with the acoustically determined respiratory flow.

Most previous works for airflow estimation relied heavily on recordings performed over the trachea. Tracheal sounds are typically very harsh and contain high frequencies easily heard during the two breathing phases [6]. This happens mostly due to the large diameter of the trachea and the absence of a structure to filter the sound [5,6]. Usually, when clinicians auscultate a patient, they perform it at several auscultation points throughout the lungs, specifically when trying to listen to adventitious respiratory sounds [6]. Therefore, the trachea region is not a common auscultation area. Besides, it can be uncomfortable for patients to have a stethoscope/microphone pushed against their throat. Another disadvantage of relying on tracheal sound is that it can be technically challenging to develop a wearable device that can record respiratory sounds in the trachea over long periods of time.

Some of the above-presented methods were also developed and tested on small datasets and healthy subjects. It is known that healthy subjects might present different breathing patterns compared to diseased subjects [26]. Thus, these methods may not apply to subjects with respiratory diseases, limiting their usefulness.

It should be noted that all previous works refer to the estimation of the respiratory airflow in terms of volume variation per unit of time. In all of them, the pneumotachograph was used to obtain the true value of the respiratory airflow.

In the context of flow–sound relation, so far, deep learning models have only been used to determine the respiratory phase, namely inspiration and expiration, using larger databases [21,27]. A model based on object detection networks (Faster R-CNN) was able to detect expiratory and inspiratory segments using the Short Fast Fourier Transform to represent the respiratory sound [5]. That method was developed with data from the Tromsø 7 lung sound dataset and achieved an average sensitivity of 97% and an average specificity of 84%. The recordings were obtained at several chest locations. In another study, a different architecture to solve the same problem, namely a hybrid deep learning model, was proposed [21]. That convolutional-recurrent model was developed using the HF_lung_V1 database and achieved an F1 score of 86.1% and 70.0% to detect inhalation and exhalation segments accordingly. The recordings were also collected at several chest locations. An extensive annotation process was required to identify the respiratory phases to train the models in both of these studies [5,21]. Such annotation processes are usually error-prone and time-consuming, thus limiting the scalability of the developed models.

Unlike the above-mentioned deep learning approaches used for respiratory phase detection, our model can estimate the complete respiratory airflow curve, providing a better overall characterization of the respiratory patterns. Our model also does not require any previous patient-specific calibration, and the data used to train our models does not require any human interaction for annotation, enhancing its scalability.

3. Materials and methods

This section describes the data used for this study and the proposed methodological framework. To process the respiratory sound, we used MATLAB 2021b. All deep learning models were developed using Python 3.8 with Tensorflow [28]. The models were trained on an NVIDIA RTX A5000 with 24 GB of GDDR6 RAM. The computer was also equipped with an Intel® Xeon(R) Silver 4214 CPU @2.20 GHz and 320 GB of RAM.

The data used in this study were collected at the Respiratory Research and Rehabilitation Laboratory, School of Health Sciences of the University of Aveiro (Lab3R-ESSUA). The study was conducted under the scope of the European Horizon 2020 project WELMO [29][1] and an independent ethics committee from the Nursing School of Coimbra (ESEnfC) approved it (Reference AD1 P721-10/2020). Informed written consent was obtained from all participants before the examinations.

Fig. 1 presents the three major steps related to the proposed methodology: (1) data preparation and augmentation (spectrogram computation and data windowing); (2) development of the deep learning models; (3) performance assessment.

3.1. Database

In this study, we used data from the BRACETS database, an open-access bimodal database containing respiratory sound and EIT [30]. A total of 50 participants (15 female, 35 male) from the database were considered in this study. A list of the considered subjects and the training and testing division can be found in the supplementary material. Subjects suffered from several respiratory conditions (26 — COPD, 17 — interstitial lung disease, 7 — asthma). Their average age was 67.4±8.9 years and mean BMI was 27.8±5.6 kg/m^2 (Male: Age 68.1±9.6 years — BMI 27.4±6.2 kg/m^2; Female: Age 65.7±7.2 years — BMI 28.7±4.1 kg/m^2). All participants were enrolled in a 12-week community-based pulmonary rehabilitation program.

Respiratory sounds were collected by placing an electronic stethoscope at four different positions (see Fig. 2). For each stethoscope placement, the corresponding EIT signal was also simultaneously recorded. Therefore, each acquisition comprises a pair of respiratory sound–EIT recording. In total, 396 acquisitions were collected, with a duration of approximately 20 s each.

Respiratory sound data were recorded using the 3M™ Littmann® Electronic Stethoscope 3200 with a sampling rate of 4000 Hz. The stethoscope was hand-held by a physiotherapist in the respective recording position, and the recording was started using the 3M Littmann StethAssist Software in an auxiliary computer. After recording the respiratory sound in every position considered, the recorded sounds were uploaded from the internal memory of the stethoscope to the auxiliary computer via Bluetooth. Every sound was then extracted from the StethAssist Software using the three available filtering models. In this study, we have used the "Extended" filtering mode.

EIT data were collected using the Goe-MF II EIT device (CareFusion, Höchberg, Germany). An array of sixteen self-adhesive electrodes (Blue Sensor, Ambu, Ballerup, Denmark) was attached to the chest circumference between the 5–6th intercostal space (xiphoid-sternal line), with another reference electrode on the abdomen. Small alternating electrical currents (5 mAmp) were delivered through adjacent pairs of electrodes in a sequential rotating process, and the remaining passive electrode pairs measured the resulting potential differences. A total of 208 voltages were measured per image frame. EIT data were acquired at a sampling rate of 33 images/second (33 Hz).

In the data collection process, two different types of acquisitions were performed. In the first type (tidal breathing + deep breathing - TbDb), subjects were requested to breathe quietly for a few seconds.

[1] https://cordis.europa.eu/project/id/825572

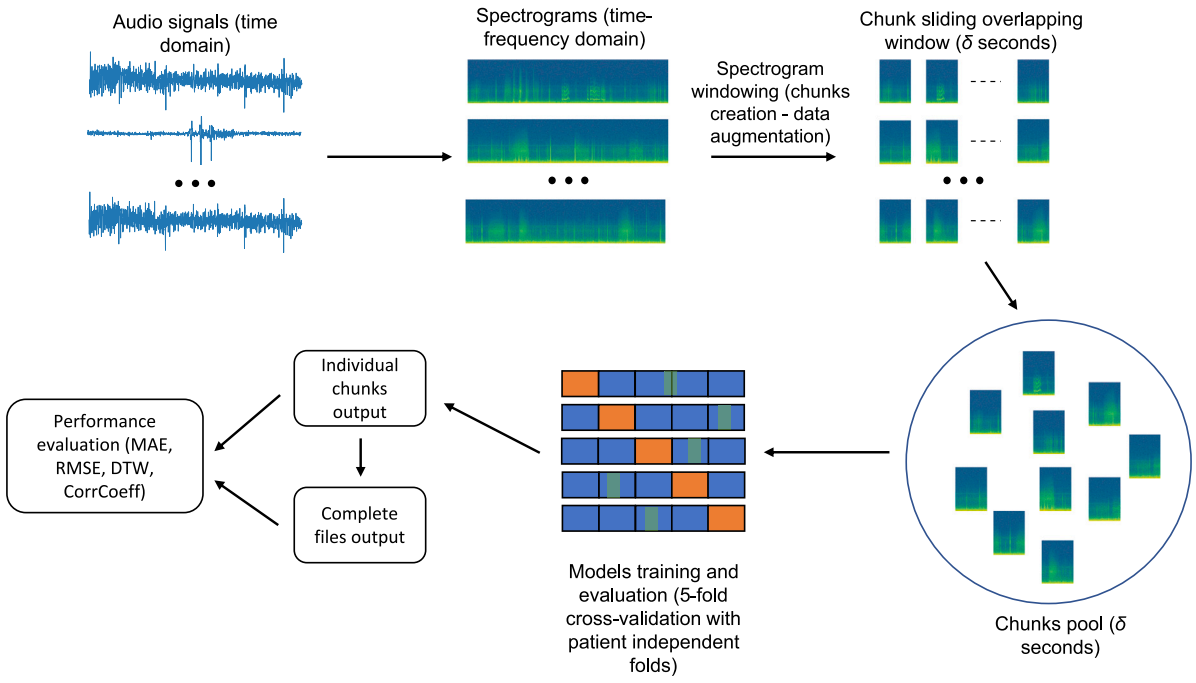


Fig. 1. Overall framework of the proposed methodology. δ — chunk time; MAE — mean absolute error; RMSE — root mean squared error; DTW — dynamic time warping; CorrCoeff — Pearson correlation coefficient.

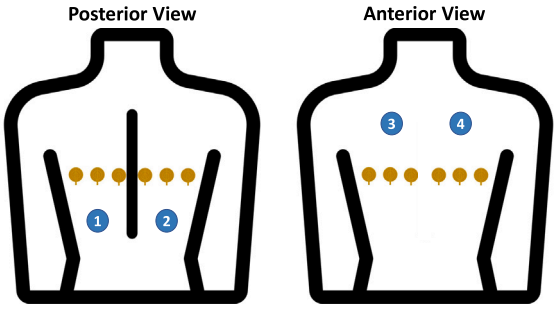


Fig. 2. Respiratory sound recording points (blue circles) and electrical impedance tomography electrodes placement (yellow circles) during examinations. 1 — posterior basal left; 2 — posterior basal right; 3 — anterior apical right; 4 — anterior apical left.

After that period, they were prompted to start breathing deeply until the end of the recording. In the second type of acquisition (tidal breathing + cough + speech - TbCS), subjects were instructed to breathe quietly at first. After a couple of breaths, they were prompted to cough (intentionally) and speak (read a sentence shown by the respiratory therapists). The sentence was in Portuguese as follows: *"Está na hora de acabar"* (in English: "It is time to end"). This sentence was selected based on a previous study where several Portuguese phrases were submitted to an extensive acoustic analysis [31]. For each recording position identified in Fig. 2, both types of acquisitions were recorded. Fig. 3 presents a subject 15-second sample of each type of acquisition.

Since we used two independent devices, the raw data from each source were not synchronized. An architecture based on an auxiliary signal was developed to synchronize respiratory sound and EIT data. The system generated an auxiliary sound signal (a pure sinusoidal tone at a frequency of 1900 Hz) that was then split to a loudspeaker and the EIT device using an audio splitter and a 3.5 mm to BNC adapter. Accordingly, this division allowed the auxiliary signal to be detected simultaneously in both respiratory sound and EIT recording systems. Subsequently, both signals were synchronized post-acquisition

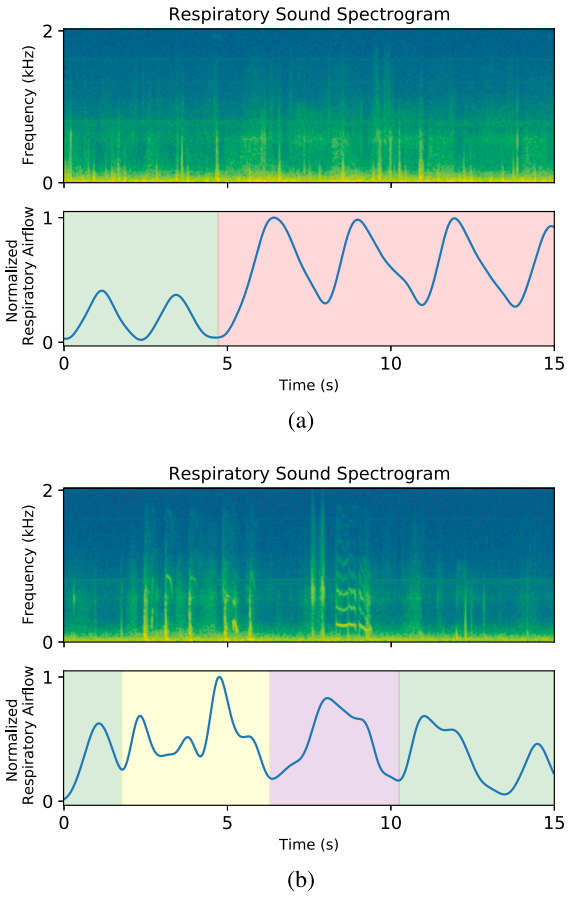


Fig. 3. Respiratory sound in one of the studied patients during two types of recording (TbDb and TbCS). (a) TbDb acquisition: tidal breathing followed by deep breathing (green — tidal breathing; red — deep breathing); (b) TbCS acquisition: tidal breathing followed by forced cough and speech (green — tidal breathing; yellow — cough; purple — speech).

by manually aligning the auxiliary signals in both sources. More details on the database are available in [30].

3.2. EIT reconstruction and ground truth creation

The dimensionless respiratory airflow information may be obtained using EIT, namely from the global EIT waveform, i.e., the sum of all reconstructed EIT images/frames pixels'. This signal has been used in multiple studies to determine breathing patterns (deep breathing, tidal breathing, among others) [17,20]. While EIT does not allow obtaining a concrete airflow value in terms of liters per minute or any other flow unit, it can measure the dimensionless respiratory airflow indirectly through the ventilation-related variation of the global impedance of the lung tissue in arbitrary units. As documented in [32], lung tissue has a resistivity which is about five times greater than most other soft tissues in the thorax. The resistivity of this tissue also increases considerably with inspiration as the alveoli stretch and electrical current has to flow around them [32]. Therefore, by monitoring this change in resistivity using EIT, the ventilation distribution can be seen over time and, subsequently, used to estimate the respiratory airflow (Fig. 3).

In chest EIT, electrodes are placed around the thorax of the patient and used for injecting electrical currents and measuring the resulting potentials through well-defined stimulation patterns. Then, using the resulting voltage measurements, reconstruction algorithms are used to obtain a 2D or 3D image of a cross-section of the lung with the respective conductivity/impedance distribution [33]. When the lungs are filling up with air, the impedance of the tissue increases. The inverse happens when they are emptying (as demonstrated in Fig. 3).

The acquired raw EIT data were processed offline to obtain the reconstructed images/frames using the Graz Consensus Reconstruction Algorithm for EIT (GREIT) [33,34]. The reconstruction was performed using an adult thorax-shaped model with a single plane of 16 electrodes. The adjacent stimulation pattern was selected from the models' library of the EIDORS software [34,35]. The resulting reconstructed EIT images consisted of 32 by 32 pixels. After obtaining the reconstructed images for every time step (frame), the global EIT waveform was computed by summing up all individual pixel values for each image. Fig. 4 represents the process of EIT reconstruction and consequent dimensionless respiratory airflow curve (global EIT waveform) computation.

After obtaining the global EIT waveform, we employed a low-pass filter with a cut-off frequency of 0.01 Hz to further smooth the curve and remove high-frequency components [20]. Because the waveform variation was expressed in arbitrary units, we have also normalized each curve between 0 and 1 (linearly scaled). When the waveform curve value was equal to 1, the lungs were at their highest impedance/resistivity value for a specific recording (typically at the end of the strongest inspiration period). Conversely, when the curve value was equal to 0, the lungs were at their lowest impedance/resistivity (typically at the end of the lowest expiration period).

Since EIT was recorded at a significantly lower sampling rate (33 Hz) in relation to the respiratory sound (4000 Hz), we have increased the number of samples of the global EIT waveform (dimensionless respiratory airflow curve) through interpolation to match the number of time-steps of the respiratory sound inputs. To do so, we used the "pcubic"[2] interpolation method. Thus, the number of time steps obtained in the representation of the respiratory sounds (see Section 3.3) was the same as the one in the respiratory airflow curve.

[2] https://www.mathworks.com/help/matlab/ref/interp1.html

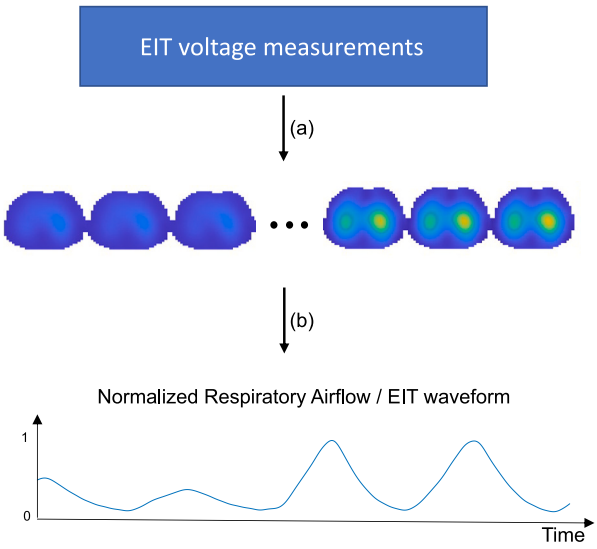


Fig. 4. Schematic representation of EIT reconstruction and consequent dimensionless respiratory airflow curve (global EIT waveform) computation: (a) EIT reconstruction using GREIT algorithm; (b) summation of all pixels for every reconstructed EIT frame. Each image presented between steps (a) and (b) represents an EIT frame (2D cross-sectional bioimpedance distribution).

3.3. Pre-processing and data preparation

We have computed the spectrograms (STFT) of the respiratory sound to use as input for the deep learning models. The spectrogram is one of the most used tools in audio analysis and processing because it describes the evolution of the frequency components over time. The STFT representation (F) of a given discrete signal is given by:

$$X(n,\omega) = \sum_{m=-\infty}^{\infty} x[m]w[n-m]e^{-j\omega m} \quad (1)$$

where $x[m]w[n-m]$ is a short-time section of $x[m]$ at time n, and $w[n]$ is a window function centered at instant n [36]. To compute the STFT, we considered a 64 ms and 128 ms Blackman–Harris window with 80% overlap [37]. For the Fast Fourier Transform (FFT), 256 points were used, resulting in 129-bin log-magnitude spectrograms. We have trained and tested the models with the inputs generated with both window sizes; however, since the 64 ms window obtained better results, we have only presented the results with that size in Section 4.

After the spectrogram computation for each respiratory audio signal, we individually normalized each spectrogram between 0 and 1. As the main objective of this work was to obtain the normalized respiratory airflow curve for each complete sound recording, performing a global normalization using all spectrograms from the different sounds would not be possible as it would disrupt the direct relationship between each sound and the corresponding respiratory airflow curve.

Given the relatively low number of available samples (Section 3.1), we used a data augmentation approach. To artificially increase the number of available samples, we divided the normalized spectrograms of the complete sounds into multiple overlapping chunks, as shown in Fig. 1. To do so, we used a sliding window of fixed size (δ seconds) across the entire spectrogram with an overlap of 99%. In total, three different lengths for the size of each chunk were considered, namely, 6 s, 10 s, and 15 s, respectively. The same windowing process was also applied to the ground truth values for the dimensionless respiratory airflow curves. It should also be noted that zero padding was applied whenever the last chunk was bigger than the audio signals spectrograms. The main objective of developing models with different-sized chunks was to understand whether the sequence length would impact

Table 1
Number of respiratory sound chunks per chunk size after data windowing (data augmentation).

Chunk size (s)	# Chunks/Samples	Input shape
6	85948	(129, 235, 1)
10	38507	(129, 392, 1)
15	14175	(129, 588, 1)

the performance of the models. Considering that the average respiratory rate typically ranges between 12 to 20 breaths per minute [38], we considered the smaller chunk size as 6 s in an attempt to have two respiratory cycles in every sample. The numbers of resulting sound chunks for each δ value after the augmentation process are presented in Table 1. The corresponding input shapes for the models for each chunk size are also presented.

3.4. Deep learning model architecture

In this work, we had the primary goal of developing a model that could modulate the dimensionless respiratory airflow over time when provided with a raw respiratory sound. Accordingly, we developed hybrid deep learning models resulting from a combination of convolution and recurrent modules (CNN + LSTM). While the convolution block automatically extracted features from the time–frequency representation of the respiratory sounds (i.e., feature learning), the recurrent module was used to learn their temporal dependencies and modulate the respiratory airflow sequence over time. Because we wanted to obtain an output for every temporal value of the input spectrogram (many-to-many regression problem), we considered 1D (one-dimension) convolutional layers. We opted for 1D convolutional layers so that they were able to learn the features at every frame and model each one individually. Then, the recurrent layers (LSTM) mapped the relationship between the learned features from each time step. Lastly, the final layer of the network (dense) was also distributed across every temporal instant. Therefore, our model produced one output for each temporal instant of the input (many-to-many relationship).

Our model was composed of two large building blocks, the convolution and the recurrent modules (Fig. 5). The convolution module consisted of 3 residual blocks. Instead of learning a direct mapping between the input of each layer and the extracted features, the residual architecture uses the difference between a mapping applied to the input and the original input. Previous studies suggested that convolutional layers were better at learning on the residual of a feature map instead of directly on the feature map [39]. The structure of the residual model is represented in Fig. 5. After the convolutional module, a 1D global maximum pooling layer was applied to concatenate the extracted features from all filters per time step, maintaining the original number of temporal instants. After the global pooling process, the samples were fed to three bidirectional LSTM layers with 256, 128, and 64 hidden units, respectively. The LSTM layers mapped the temporal dependencies between learned features of the temporal instants. Lastly, we used a time-distributed dense/fully connected layer with a linear activation. The last layer mapped the input to the corresponding normalized respiratory airflow and estimated the final frame-wise prediction.

We performed grid-search experiments using filters with size 3 and filters with size 5 and verified that the results were better using size 5. Therefore, all convolutional layers comprise filters with size 5. We have also tested the architecture with unidirectional LSTM layers, but we have verified that the output of the models was much noisier when compared to bidirectional layers. Thus, we kept the bidirectional layers. Besides those parameters, we also experimented with a different number of convolutional and recurrent layers and kept the configuration with better results in our preliminary testing. In total, the final model had 2,023,425 parameters.

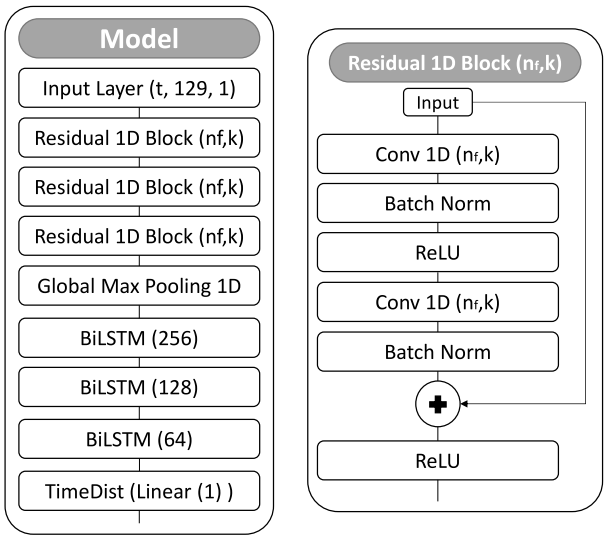


Fig. 5. Block diagram representation of the architecture of the deep learning model. t — number of time steps for each chunks size (see Table 1); n_f = 128 (number of convolutional filters); k = 5 (kernel size); TimeDist — time distributed layer.

Table 2
Hyperparameters used for the training of the deep learning models.

Hyperparameter	Value
Dataset Partition	5-fold cross-validation (Subject isolation)
Learning Rate	3e−4
Number of Epochs	100
Loss Function	RMSE
Optimizer	ADAM

The network models were trained for 100 epochs. Simultaneously with the training process, the model was evaluated using the validation subset at every new epoch to save only the set of weights with the lowest validation loss. Table 2 presents the parameterization used in the models' training process. It also presents the used data split, which will be further explained in Section 3.6.

3.5. Post-processing

As previously mentioned, we split the complete respiratory sounds into multiple fixed-size chunks using an overlapping sliding-window approach (see Section 3.3). Therefore, the developed models were trained using these chunks; consequently, their output was also a chunk-wise prediction for the respiratory airflow. To obtain the respiratory airflow sequence for the complete audio recordings when testing the models, we used a sliding ensemble method [40]. Using this approach, we combined the predicted value for the respiratory airflow for each time frame. We averaged that value across all chunks to obtain the final prediction for the whole sound signal. We only averaged the predicted airflow for the overlapping indices/time-frames because we used partially overlapping sliding windows. Fig. 6 presents an example of the process of chunk grouping. Then, we applied a 5th-order Butterworth filter with a cutoff frequency of 0.05 Hz to smooth any high-frequency variations in the predicted curves. Lastly, we normalized the complete sound output between 0 and 1 to ensure the output stayed within this interval. In the supplementary material, an example can be found with the post-processing of a recording.

3.6. Evaluation tasks and metrics — internal validation

We created four evaluation tasks based on the type of recording described in Section 3.1 to understand how the models behave under

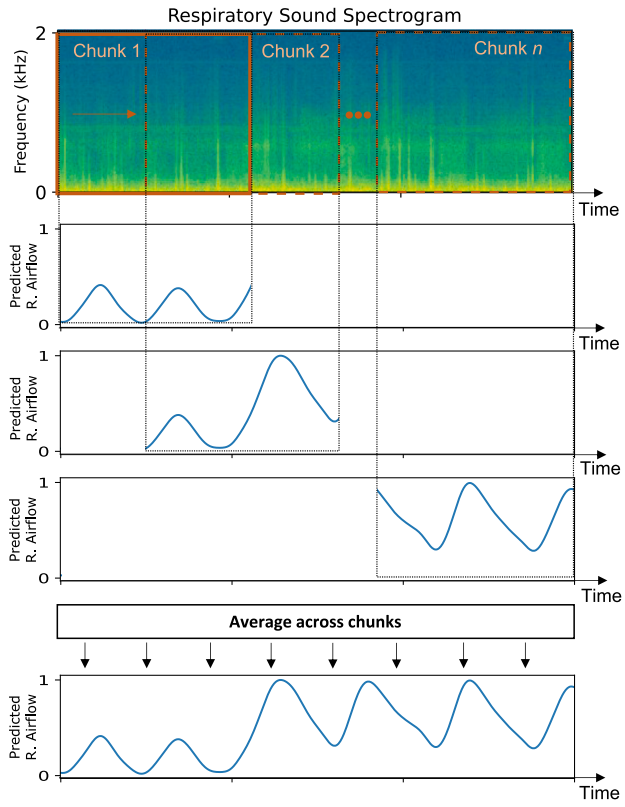


Fig. 6. Example of the output conversion from each chunk to the complete respiratory sound. The predicted sequences for the respiratory airflow of each chunk were combined into a single probability respiratory airflow sequence by averaging the obtained values for all individual chunks in overlaying frames.

Table 3
Division of file type for each evaluation task training and testing sets. TbDb - tidal breathing + deep breathing; TbCS - tidal breathing + cough + speech (Section 3.1 for file type description).

	Train	Test
Task I	TbDb	TbDb
Task II	TbDb + TbCS	TbDb + TbCS
Task III	TbDb + TbCS	TbDb
Task IV	TbDb	TbDb + TbCS

different circumstances and how they can generalize. In *Task I*, we investigated whether the networks could model the respiratory airflow using sound without external perturbations, such as speech and cough. *Task II* represents the closest to the real-world task because all file types were considered for training and testing, respectively. In *Task III*, we wanted to understand if the networks could better model the respiratory airflow in files without perturbations by using all files for the training process. Finally, in *Task IV*, we wanted to understand how the networks trained only with files without perturbation would behave when used to model the respiratory airflow in all file types. Table 3 summarizes the type of files used for training and testing the models in the different tasks. Given that the patient split (described below) was the same across tasks, we used the models trained in *Task I* and *Task II* for testing in *Task IV* and *Task III*, respectively, to reduce the computational load. Thus, in *Tasks III* and *IV*, only the files in the test set change.

In order to obtain a reliable estimate of the performance of the models in all different tasks, we performed a 5-fold patient-independent cross-validation scheme. Thus, 40 subjects were used in the training set in each fold, and 10 were left for the testing set. Moreover, in each fold, we randomly selected 12.5% of the tanning subjects (5 subjects) for validation purposes. With this strategy, every subject belonged exclusively to either the training or testing set of each fold (no data leakage). Additionally, we used every subject to evaluate the model because everyone was in the test set once. Typically, samples from the same patient tend to have some similarity within themselves, which might lead to overly optimistic performance results whenever data from the same subject is in both sets [34]. Additionally, the main objective for real-world applications is usually to deploy the models in new subjects, stressing the need for patient-independent validation.

Several statistical metrics commonly used in regression problems were considered to evaluate the trained models. We have used the mean absolute error (MAE), rooted mean squared error (RMSE), dynamic time warping similarity (DTW) [41],[3] and Pearson correlation coefficient value (PCC). While MAE and RMSE were used to analyze the overall fit and error of the predicted values compared with the true values, DTW and PCC were applied to evaluate whether the morphology of the respiratory airflow curve followed a similar trend compared to the original one. The DTW measures the similarity between two temporal sequences and calculates their distance. On the other hand, PCC was used to assess how similar the trend was between the original and predicted curves (that is, if the curves had a similar evolution over time in terms of their trend). The equations for each considered evaluation metric are presented below:

$$MAE(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2}$$

$$RMSE(y,\hat{y}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3}$$

$$PCC(y,\hat{y}) = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\cdot\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}} \tag{4}$$

$$DTW(y,\hat{y}) = \min_{\pi}\sqrt{\sum_{(i,j)\in\pi} d(y_i,\hat{y}_i)^2} \tag{5}$$

where, y is the original sequence, $\hat{y}$ the predicted sequence, and π is a path from the cross-similarity matrix obtain from y and $\hat{y}$.

3.7. Respiratory cycle detection — external assessment

Since we have designed a method based on an ensemble approach, our models can deal with different sound durations. Therefore, we have used an external database to understand better how our models would perform when tested with respiratory sounds other than those considered in this study scope. The Respiratory Sound Database (RSD) [42,43] was used as an independent external assessment set. This is the only database with a complete annotation of respiratory cycles that is freely available. The RSD contains audio samples collected independently by two research teams in two countries. It is a challenging database as the recordings contain several types of background noises and sounds. The sounds were collected using multiple acquisition systems (AKG C417L Microphone (AKGC417L), 3M Littmann Classic II SE Stethoscope (LittmannC2SE), 3M Littmann 3200 Electronic Stethoscope (Littmann3200), WelchAllyn Meditron Master Elite Electronic Stethoscope (Meditron)). It is also worth noting that the RSD acquisition protocol differed from the one in our study. We have not considered RSD files that were recorded at the tracheal site.

To the best of our knowledge, there are currently no freely available respiratory sound databases with respiratory airflow information or respiratory sound and EIT. While the RSD does not have this type of information, it has the annotations of complete respiratory cycles composed of (inspiration and expiration). Therefore, using this database, we could only assess the models regarding their ability to correctly

[3] https://tslearn.readthedocs.io/en/stable/user_guide/dtw.html

Table 4
Qualitative evaluation scores description.

Score	Description
5	Excellent correspondence between estimated and annotated respiratory cycles. Respiratory airflow curve with good overall shape and behavior and excellent correlation with respiratory phases.
4	Good correspondence between estimated and annotated respiratory cycles. Respiratory airflow curve with good overall shape and behavior.
3	Average correspondence between estimated and annotated respiratory cycles. Respiratory airflow curve with average overall shape and behavior.
2	Few identifiable matches between estimated and annotated respiratory cycles. Respiratory airflow curve with irregular behavior.
1	No identifiable matches between estimated and annotated respiratory cycles. Respiratory airflow curve with irregular patterns.

identify the respiratory cycles and the overall behavior of the output curves, not the estimation of the relative respiratory airflow of each recording in terms of amplitude. Even though this data did not allow us to completely validate our model, we could assess their performance and behavior with data from different patients, recording devices, and recording protocols. Therefore, it served as an external assessment element to better understand the generalization capabilities of the models under different circumstances.

We have performed a visual qualitative assessment to evaluate how the identified respiratory cycles from the respiratory airflow curve aligned with the annotated respiratory cycles, aiming at validating our models in the RSD. We visually inspected how well the lower inflection points of the estimated airflow curves matched the annotated respiratory cycles. Based on the alignment, we have given each estimated respiratory airflow curve a score according to the criteria defined in Table 4.

Fig. 7 presents two examples of the estimated respiratory airflow curve for two RSD files ("185_1b1_Pl_sc_Litt3200.wav" and "205_1b3_Pr_mc_AKGC417L.wav"). The black dotted vertical lines represent the annotated respiratory cycles. In Fig. 7(a), we can see a near-perfect match between all annotated respiratory cycles and the local minima inflection points of the estimated respiratory airflow curve, meaning that the model correctly identifies the cycles. Moreover, the estimated curve presents a regular behavior with excellent overall shape. Thus, the estimation for this specific respiratory sound would score a 5. On the other hand, in Fig. 7(b), we cannot see any match between the annotated respiratory cycles and the inflection points of the estimated respiratory flow. Moreover, the estimated curve presents an irregular behavior with no discernible breathing patterns. Therefore, this specific respiratory sound would score a 1. In the supplementary material, a document can be found with an example for all scores (from 1 to 5).

To perform this qualitative analysis, we have considered all the 60 files from the RSD recorded using the Littmann 3200 Stethoscope (the same device was used in our study) and another 60 randomly selected files recorded with the other devices (20 from each recording device). The qualitative analysis was performed by the first author (who has extensive experience in analyzing both EIT and respiratory sound signals).

4. Results

As previously discussed in Section 3, we have considered two types of evaluation. In the following subsections, we present the obtained results divided by the type of evaluation, respectively.

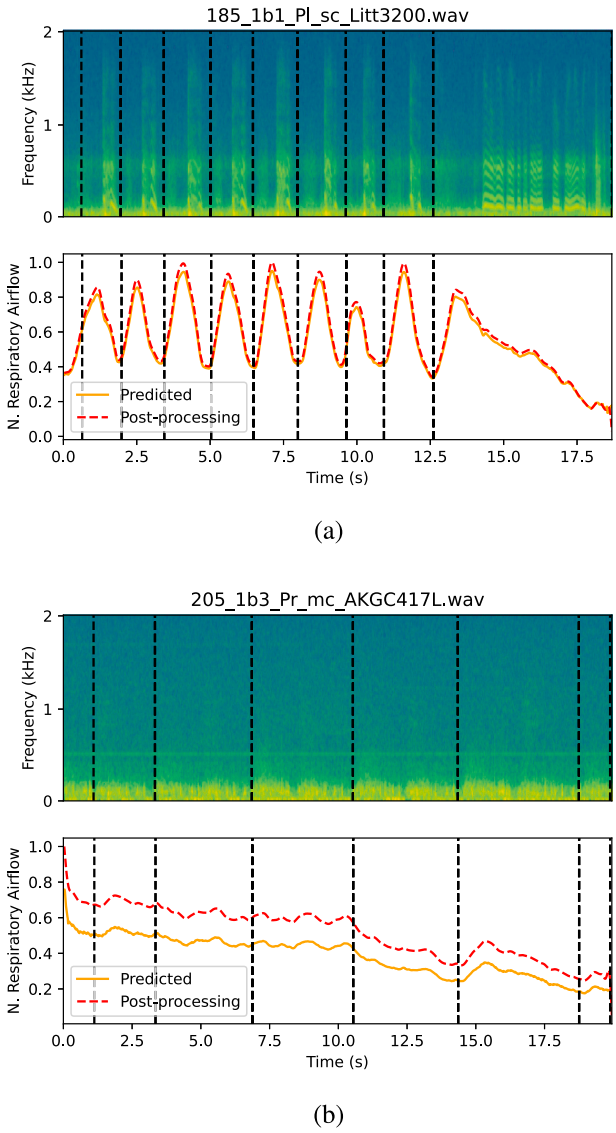


Fig. 7. Example of normalized respiratory airflow estimation for RSD files. (a) RSD file: 185_1b1_Pl_sc_Litt3200.wav (Score - 5); (b) RSD file: 205_1b3_Pr_mc_AKGC417L.wav (Score - 1).

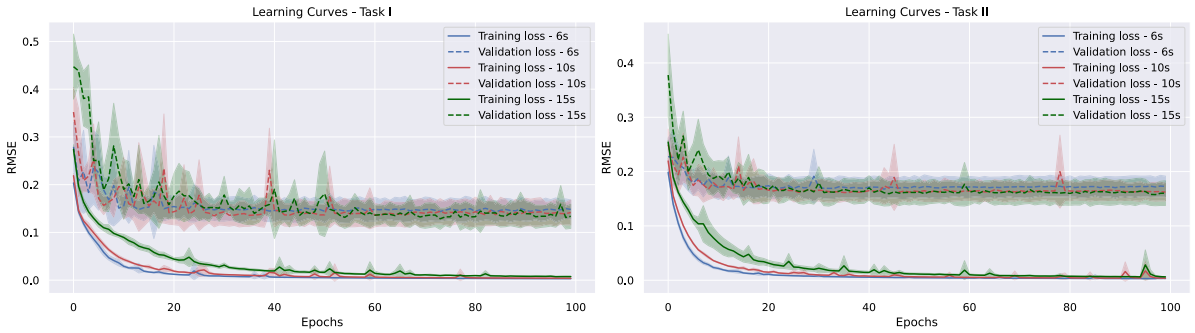
4.1. Internal validation

To assess the performance of the developed models, we have considered four different evaluation tasks of different complexity (Table 3). In each of them, we varied the set of files that were used, both in the training and testing sets, as described in Section 3.6. We have also computed the results for the output of the models when considering the individual chunks output and the complete files output (merging process described in Section 3.5). Table 5 presents the obtained results in the test set for all considered evaluation tasks. The results were obtained using a 64 ms window to compute the spectrograms (see Section 3.3). We have also plotted the learning curves for the models across all folds in Fig. 8.

In *Task I*, the models developed using 15-second chunks were globally the best performers, with better results in almost all considered metrics, both for the chunks and complete files output, respectively. On the other hand, models developed using the smaller 6-second segments had the worst performance. In *Task III*, nearly identical results were obtained. Here, the models developed using 15-second chunks were also the best, and the ones using 6-second chunks were the worst. In

Table 5
Results summary table separated by task, with the mean value across all folds (mean ± standard deviation). The best results per metric within each task were highlighted in bold for the chunks and complete files.

Task	Chunk time (s)	Type of sample	MAE	RMSE	DTW	CorrCoef
Task I	6	Chunks	0.118 ± 0.084	0.144 ± 0.097	**1.982 ± 1.871**	0.848 ± 0.272
		Complete Files	0.112 ± 0.056	0.144 ± 0.068	3.024 ± 1.786	0.877 ± 0.161
	10	Chunks	0.110 ± 0.064	0.138 ± 0.077	2.141 ± 1.666	**0.890 ± 0.183**
		Complete Files	0.114 ± 0.055	0.146 ± 0.068	2.990 ± 1.626	0.875 ± 0.164
	15	Chunks	**0.105 ± 0.052**	**0.135 ± 0.070**	2.435 ± 1.581	0.886 ± 0.166
		Complete Files	**0.110 ± 0.051**	**0.142 ± 0.067**	**2.859 ± 1.585**	**0.880 ± 0.157**
Task II	6	Chunks	0.144 ± 0.083	0.171 ± 0.094	**2.457 ± 1.806**	0.780 ± 0.304
		Complete Files	0.143 ± 0.059	0.180 ± 0.072	3.841 ± 1.809	0.769 ± 0.212
	10	Chunks	0.135 ± 0.068	0.166 ± 0.080	2.709 ± 1.713	**0.793 ± 0.251**
		Complete Files	0.138 ± 0.058	0.174 ± 0.071	3.553 ± 1.685	**0.770 ± 0.221**
	15	Chunks	**0.129 ± 0.060**	**0.161 ± 0.074**	2.934 ± 1.567	0.786 ± 0.237
		Complete Files	**0.134 ± 0.061**	**0.170 ± 0.075**	**3.282 ± 1.514**	0.770 ± 0.235
Task III	6	Chunks	0.125 ± 0.085	0.151 ± 0.098	**2.132 ± 1.888**	0.841 ± 0.291
		Complete Files	0.118 ± 0.057	0.152 ± 0.072	3.290 ± 1.847	0.858 ± 0.187
	10	Chunks	0.112 ± 0.065	0.140 ± 0.079	2.182 ± 1.607	**0.881 ± 0.219**
		Complete Files	0.113 ± 0.054	0.145 ± 0.069	2.990 ± 1.628	**0.869 ± 0.191**
	15	Chunks	**0.106 ± 0.060**	**0.134 ± 0.076**	2.362 ± 1.487	0.877 ± 0.223
		Complete Files	**0.109 ± 0.057**	**0.141 ± 0.074**	**2.679 ± 1.421**	0.869 ± 0.216
Task IV	6	Chunks	0.181 ± 0.116	**0.219 ± 0.133**	**3.165 ± 2.412**	**0.582 ± 0.469**
		Complete Files	**0.169 ± 0.080**	**0.215 ± 0.097**	3.894 ± 2.030	**0.610 ± 0.353**
	10	Chunks	0.181 ± 0.100	0.224 ± 0.118	3.370 ± 2.188	0.554 ± 0.444
		Complete Files	0.174 ± 0.081	0.222 ± 0.099	**3.803 ± 1.772**	0.584 ± 0.370
	15	Chunks	**0.179 ± 0.094**	0.226 ± 0.115	3.719 ± 2.158	0.532 ± 0.430
		Complete Files	0.178 ± 0.086	0.227 ± 0.106	3.927 ± 1.940	0.555 ± 0.396



(a) (b)

Fig. 8. Graphical representation of the mean and standard deviation of training and validation learning curves obtained by averaging all five developed models in each fold with the different chunk sizes. (a) Learning curves Task I (same for Task IV); (b) Learning curves Task II (same for Task III).

Task II, where all file types were considered, the performance decreased due to the increasing complexity of the respiratory airflow curves in respiratory files with perturbations, such as cough and speech. Once again, in this task, the models developed with the larger chunks were the ones that globally performed the best. However, the models using 10-second inputs obtained slightly better results regarding the correlation coefficient value. Lastly, *Task IV* was, by a significant margin, the task where the models struggled the most, with the lowest performance results compared to the remaining tasks. In this case, the chunk size had the reverse impact on the results, as the values obtained from 6-second models presented overall better results.

4.2. External assessment

We analyzed the mean value for the Pearson coefficient correlation value for all runs and configurations considered (6, 10, and 15 s) in *Task II* and chose the model with the highest value for this specific metric to assess how our model would perform when tested on external data. Then, using the chosen model, we tested it on the RSD and gave a score to every estimated respiratory airflow curve based on a visual qualitative assessment, as described in Section 3.7 and Fig. 7. The results of this analysis are depicted in Table 6, where the percentages of files with each score (grouped by recording device) are presented.

The results presented in Table 6 showed that the model performed better on external data that were recorded using the same device as in the current study (Littmann 3200). This device had the highest average score, with more than 80% of the files scoring a 4 or a 5. We also observed that, with this device, more than 97% of the recordings scored 3 or more, with a residual percentage of 3% scoring a 2. For the Littmann C2SE, the obtained average score was almost the same, with a larger standard deviation. However, a higher percentage of estimations were classified as a 5 for this device. When considering the remaining devices, the performance decreased substantially, with more than 50% of the files recorded with AKGC417L microphone and Meditron digital stethoscope scoring below 3.

5. Discussion

When globally analyzing the results obtained across all tasks in Section 4.1, we observed that the models performed better in *Task I*

Table 6
Qualitative analysis results (percentage of files with each score, divided by equipment).

Equipment \Score	5	4	3	2	1	Avg.
Littmann 3200	53%	28%	15%	3%	0%	4.3 ± 1.1
Littmann C2SE	75%	5%	10%	10%	0%	4.3 ± 1.4
AKGC417L	30%	5%	5%	30%	30%	2.2 ± 1.2
Meditron	25%	5%	15%	55%	0%	2.0 ± 1.5
Global	48%	17%	13%	18%	5%	3.6 ± 1.3

and *Task III*. The results were nearly identical in both tasks. Such results were previously expected, as the testing set files in both tasks did not contain systematic perturbations such as cough and speech. Another interesting observation was related to the worst results obtained in *Task IV*. In this task, not only were the results significantly lower, but they were also much more dispersed, as seen by the larger standard deviations. Since only TbDb files were used for the training process and both types were used for the testing, we can infer that training the deep learning models with different types of files was beneficial. Using both file types can improve the overall generalization capability of the model.

From the analysis of the results in Section 4.1, one pattern was noticed in *Tasks I*, *II*, and *III*: the models developed considering 15-second chunks were the ones with the best overall average performance in most evaluation metrics, both at chunk and complete files level. Considering this, we hypothesized that the larger sequence sizes, which gave the model more contextual information, helped to achieve a better estimation of the respiratory airflow. Moreover, we have also observed that, in general, the models using bigger chunks produced a better output in terms of the amplitude of the predicted curves. On the other side of the spectrum, the smaller sequence sizes, with fewer respiratory cycles, led the models to perform the worst, providing further evidence that more contextual information leads to better estimations. This aspect is mostly related to the recurrent module (see Section 3.4). This pattern presented an opposite behavior for *Task IV*, in which several metrics yielded better results when considering smaller chunks. However, the difference between chunk sizes was not as distinct as in the remaining tasks, and the standard deviations were larger.

One of the main reasons for the lower results in tasks that considered files with perturbations (TbCS files) was related to their inherent higher complexity. For instance, while some subjects took a deep breath and then spoke continuously, others started speaking and then took a breath in between words. Similarly, such variations were also observed during the coughing stage. For instance, while some subjects took a deep breath and coughed, others started coughing and inhaled before coughing again. Also, after coughing, some subjects would experience dyspnea, which, in consequence, led to sudden gasps for air. Such variations during the cough and speaking phases introduced considerable variability in the airflow waveform samples. Therefore, each sample is very specific to a particular subject, and there was substantial heterogeneity in the airflow curve patterns, both within the same subject and different subjects. This heterogeneity between different subjects was also verified for TbDp files, albeit at a much lower degree.

Two examples of model outputs for two complete files, one for each type of acquisition, are presented in Fig. 9. In Fig. 9(a), we can observe that the model can estimate the respiratory airflow generally well for the TbDb acquisition, not only in terms of amplitude but also in terms of behavior and morphology. Regarding the example in Fig. 9(b), the model can also estimate the respiratory airflow curve with a high degree of fidelity. In this concrete example, we see that the overall morphology and behavior of the estimated curve are identical to the original. However, it presents a slight mismatch in amplitude for the respiratory cycles with cough. This behavior for the estimated curves of both file types was broadly seen throughout all the performed tasks. When TbDb files were considered, the amplitudes of the estimated curves were closer to the original ones. Despite that, there were still

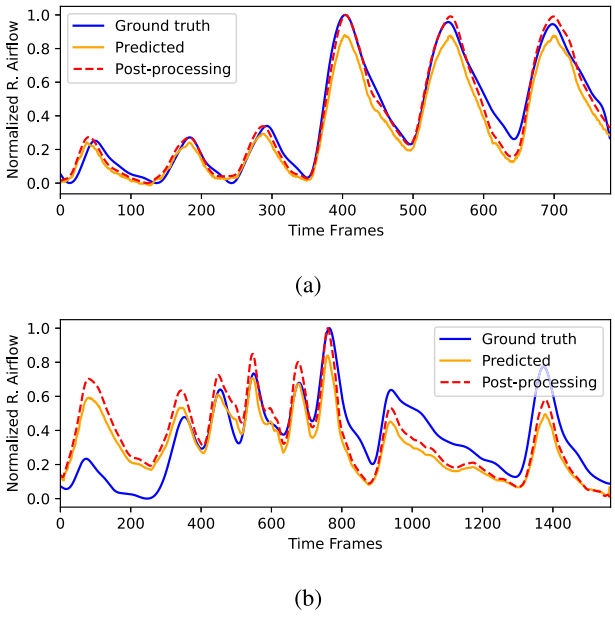


Fig. 9. Example of dimensionless normalized respiratory airflow output for complete files. (a) Example of model output for TbDb file; (b) Example of model output for TbCS file.

several cases where the estimated curves were very different in terms of amplitude and behavior when compared to the ground truth.

In general, we found out that our models struggled in TbDb files when the respiratory sounds had intense background noise that presented a broad spectral signature across the spectrograms' frequency axis. Moreover, other situations that we observed to be particularly difficult for the models were when the breathing patterns were very shallow, making it very hard to visually identify any discernible patterns related to the breathing process in the spectrograms. In both cases, a human annotator would also struggle to identify/hear the breathing phases of the respiratory sound. On the other hand, in TbCS files, the models typically struggled the most whenever the power of the perturbations was higher in the time–frequency representations. Typically, this led to an amplitude misestimation of the dimensionless respiratory airflow curves.

In Fig. 8, we observe a gap between the learning curves of the training and validation losses for the models with all three chunk sizes. This gap may occur if the training dataset has too few examples compared to the validation dataset (unrepresentative training set). Nevertheless, our models did not overfit during training, as in both *Tasks I* and *II*, the average training and validation losses show improvement over time and did not increase. The number of subjects in the training dataset can be increased to address the gap between curves. However, we decided to keep a bigger number of subjects in the validation set (five subjects) to avoid overfitting. Another interesting remark is that the bigger the size of the chunks, the slower the learning curves converged. This was mainly related to the tuning of the parameters of the recurrent layers for the bigger chunk sizes because they had more temporal instants. In Fig. 8, we also observed that the mean validation loss for the 15 s models was generally lower than the other sizes.

When considering the external assessment process (Table 4), from the analysis of the different spectrograms obtained from the different recording devices, we have found that their morphology was quite contrasting. Therefore, we hypothesize that the decrease in performance observed when the model was applied to respiratory sounds recorded with different devices was mainly related to the morphological differences between spectrograms obtained with the Littmann 3200 and the other devices. For instance, in Fig. 7, the morphological

differences between spectrograms of different devices are contrasting, with a very different spectral signature. While the sound recording with the AKGC417L does not have any processing, the sound recorded with the Littmann 3200 is filtered by the device. This leads to a different energy distribution across the frequency axis. Given that the models were only trained with a particular device in this study, they behave better when considering that same device. From the analysis of Table 4, we also observed that the Littmann C2SE performed well, with a significant percentage of the files with the maximum score. This was mostly related to the similar characteristics of the spectrograms compared to the Littmann 3200. However, it is also worth noting that most of the Littmann C2SE files were recordings of subjects in tidal breathing with little noise. Those files were similar to the TbDb considered in our study.

Another interesting observation made when performing the external qualitative analysis on the RSD was that, in some cases, the inflection points of the respiratory airflow curves were slightly offset compared to the respiratory cycle annotation. However, upon closer inspection, we concluded that, in fact, in those cases, the respiratory cycles were marginally mislabeled. Although we have not conducted any extensive validation, this fact leads us to believe that our method can be more precise/sensitive when compared to human annotators, providing more accurate detection and according delimitation of the respiratory cycles. Some examples of those instances can be found in the supplementary material. To better characterize the distance between the inflection points of the estimated airflow curves and the annotated respiratory cycles, we have manually annotated the inflection points of all estimated airflow curves with a score equal to 3 or superior. Then, we computed the time distance between the inflection points and the real respiratory cycles: Score = 3 — difference = 0.27±0.23 s; Score = 4 — difference = 0.19±0.19 s; Score = 5 — difference = 0.18±0.16 s. As expected, the higher the score, the closer the inflection points of the estimated curves were to the annotations of the respiratory cycles.

Regarding computational complexity, we timed how long it took to run on the external RSD files. For files with approximately 20 s, the complete estimation of the dimensionless respiratory airflow curve, respective post-processing, took about 4 s. These files were tested on the same machine described in Section 3 to train the models. Despite a high running time for deployment in a real-time application, our method is still suitable for implementation in a wearable device with off-device processing. For instance, a wearable device could record respiratory sounds and retrieve them for a cloud-based server for algorithmic processing and analysis (similar architecture to [15]).

To the best of our knowledge, this was the first work where a deep learning-based approach had been used to estimate the respiratory airflow signal solely based on the respiratory sound. Even though our models performed relatively well across all evaluation tasks, except for *Task IV*, there is significant room for improvement. From our external assessment, we have learned that our model performs well when used on external data recorded with the same device used to record the respiratory sounds in this work, namely the Littmann 3200. We also verified that in the files from the Littmann C2SE device, with few internal and external perturbations, the model also performed well in the qualitative analysis. However, we observed a significant drop in performance when considering other devices, implying that the model does not have enough generalization capacity to cope with sounds from those devices. As happens in all machine learning works, the diversity of the datasets used for training is a factor that can highly affect the performance of the models. Because our database only contained data from one specific device, we expected our models to be better suited for samples acquired with that same device. In order to overcome this limitation of our models, we should integrate data from other devices acquired under different circumstances and protocols to maximize their generalization capability. In addition, methods to normalize the spectral sensitivities of the recording devices should also be explored.

Ultimately, one of the greatest potentials of our methodology is that it can be paired with other automated methods and pave the way for the development of more advanced and autonomous pipelines in the area of automated monitoring and processing of respiratory sounds.

For instance, the method could be used to extract the dimensionless respiratory airflow from a respiratory sound recording and, on top of it, use another method to extract spirometric parameters from the airflow curve. Previous studies in the literature have used the EIT waveform to extract spirometric parameters such as IVC, FEV1, FEV1/FVC, among others [44–46]. With such an approach, extracting these measures based on solely respiratory sound would be possible (which, in comparison to classical spirometry, would be less dependent on patient cooperation and easier to deploy in remote settings with wearable devices; moreover, unlike spirometry, it would not require patients to wear a nose clip and mouthpiece). The spirometric indices are very well established for normality and disease cases and useful for the differential diagnosis of respiratory diseases. Moreover, they are some of the most common indicators used by clinicians to diagnose respiratory diseases and monitor their evolution over time.

Similarly, after the extraction of the dimensionless airflow curve, methods to perform breathing phase detection [20] and adventitious respiratory sounds segmentation (applied to the respiratory sounds [47]) could be employed. Then, the timing of these adventitious sounds could be determined (e.g., late expiratory, early inspiratory). Previous studies highlighted the clinical relevancy of the timing of adventitious sounds and correlated them with several respiratory diseases [5,11].

6. Conclusion

The advent of electronic stethoscopes/microphones, coupled with significant progress in electronics, machine learning, and signal processing, has revolutionized auscultation, as computer-assisted decision/monitoring systems are now becoming increasingly more common. The method developed in this study to estimate the dimensionless respiratory airflow aims to enhance further and promote the implementation of computer-based approaches to facilitate the development of more advanced and autonomous pipelines in the automated monitoring and processing of respiratory sounds. Such pipelines could be particularly beneficial in telemonitoring/telehealth applications when implemented with wearable devices.

Our study has demonstrated that obtaining the dimensionless respiratory airflow signal is feasible using only the respiratory audio signal and deep learning models. Nevertheless, some of the main limitations of our models were related to their generalization capacity, highlighted when considering external data. Future work should target the collection of more data, namely using different acquisition protocols and recording equipment, to further increase the generalization capacity of the models. Other network architectures should also be explored using different representations for the audio recordings.

One of the most promising applications of our method would be its pairing with other computerized methods for respiratory sound analysis and processing, especially methods that can automatically segment adventitious respiratory sounds. With those two methodologies combined, we could provide a better overall characterization of the respiratory sound and, thus, enhance the diagnostics and monitoring capability of computerized methods for patients suffering from respiratory diseases. Therefore, significant future efforts should aim to develop further and refine computational methods to analyze respiratory sound. Also, with our method, one could envision predicting spirometric parameters directly from the dimensionless airflow curve extracted directly from the respiratory sound signal.

CRediT authorship contribution statement

Diogo Pessoa: Conceptualization, Methodology, Software, Visualization, Validation, Definition of the acquisition protocol, Data collection and organization, Writing - original draft & review. **Bruno Machado Rocha:** Conceptualization, Definition of the acquisition protocol, Data collection, Writing - review & editing. **Maria Gomes:** Definition of the acquisition protocol, Participants recruitment, Data collection, Writing - review & editing. **Guilherme Rodrigues:** Definition of the acquisition protocol, Participants recruitment, Data collection, Writing - review & editing. **Georgios Petmezas:** Data collection, Writing - review & editing. **Grigorios-Aris Cheimariotis:** Data collection, Writing - review & editing. **Nicos Maglaveras:** Definition of the acquisition protocol, Writing - review & editing. **Alda Marques:** Definition of the acquisition protocol, Participants recruitment, Writing - review & editing. **Inéz Frerichs:** Supervision, Conceptualization, Definition of the acquisition protocol, Writing - review & editing. **Paulo de Carvalho:** Supervision, Conceptualization, Funding acquisition, Definition of the acquisition protocol, Writing - review & editing. **Rui Pedro Paiva:** Supervision, Conceptualization, Funding acquisition, Definition of the acquisition protocol, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data used in this study is available at https://data.mendeley.com/datasets/f43c7snks5/1. The models developed in this study are available for public use at https://github.com/DiogoMPessoa/Dimensionless-Respiratory-Airflow-Estimation.

Acknowledgments

This work is funded by the FCT - Foundation for Science and Technology, I.P./MCTES through national funds (PIDDAC), within the scope of CISUC R&D Unit - UIDB/00326/2020 or project code UIDP/00326/2020, by FCT Ph.D. scholarships (DFA/BD/4927/2020 and SFRH/BD/135686/2018), and by the Horizon 2020 Framework Programme of the European Union project WELMO (under grant agreement number 825572).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.bspc.2023.105451.

References

[1] World Health Organization (WHO), The top 10 causes of death, 2022, (Accessed 02 Feb 2023).
[2] G.J. Gibson, R. Loddenkemper, B. Lundbäck, Y. Sibille, Respiratory health and disease in Europe: The new European Lung White Book, Eur. Respir. J. 42 (3) (2013) 559–563.
[3] GOLD, 2023 GOLD reports - global initiative for chronic obstructive lung disease - GOLD, 2023, (Accessed 02 Feb 2023).
[4] A. Marques, A. Oliveira, C. Jacome, Computerized adventitious respiratory sounds as outcome measures for respiratory therapy: A systematic review, Respir. Care 59 (5) (2014) 765–776.
[5] C. Jácome, J. Ravn, E. Holsbø, J.C. Aviles-Solis, H. Melbye, L. Ailo Bongo, Convolutional neural network for breathing phase detection in lung sounds, Sensors 19 (8) (2019) 1798.
[6] A. Bohadana, G. Izbicki, S.S. Kraman, Fundamentals of lung auscultation, N. Engl. J. Med. 370 (8) (2014) 744–751.
[7] A. Marques, C. Jácome, Future prospects for respiratory sound research, in: Breath Sounds, Springer International Publishing, Cham, 2018, pp. 291–304.
[8] A. Marques, A. Oliveira, Normal versus adventitious respiratory sounds, in: Breath Sounds, Springer International Publishing, Cham, 2018, pp. 181–206.
[9] S. Reichert, R. Gass, C. Brandt, E. Andrès, Analysis of respiratory sounds: State of the art, Clin. med. Circul. Respirat. Pulmonary Med. 2 (2008) CCRPM.S530.
[10] P. Piirilä, A. Sovijärvi, Crackles: recording, analysis and clinical significance, Eur. Respir. J. 8 (12) (1995) 2139–2148.
[11] H. Melbye, J.C. Aviles Solis, C. Jácome, H. Pasterkamp, Inspiratory crackles—early and late—revisited: identifying COPD by crackle characteristics, BMJ Open Respirat. Res. 8 (1) (2021) e000852.
[12] E. Messner, M. Hagmuller, P. Swatek, F.-M. Smolle-Juttner, F. Pernkopf, Respiratory airflow estimation from lung sounds based on regression, in: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2017, pp. 1123–1127.
[13] P.D. Muthusamy, K. Sundaraj, N. Abd Manap, Computerized acoustical techniques for respiratory flow-sound analysis: a systematic review, Artif. Intell. Rev. 53 (5) (2020) 3501–3574.
[14] S. Huq, Z. Moussavi, Acoustic breath-phase detection using tracheal breath sounds, Med. Biol. Eng. Comput. 50 (3) (2012) 297–308.
[15] V. Kilintzis, N. Beredimas, E. Kaimakamis, L. Stefanopoulos, E. Chatzis, E. Jahaj, M. Bitzani, A. Kotanidou, A.K. Katsaggelos, N. Maglaveras, CoCross: An ICT platform enabling monitoring recording and fusion of clinical information chest sounds and imaging of COVID-19 ICU patients, Healthcare (Switzerland) 10 (2) (2022) 1–19.
[16] T.A. Khan, S.H. Ling, Review on electrical impedance tomography: Artificial intelligence methods and its applications, Algorithms 12 (5) (2019) 88.
[17] I. Frerichs, M.B.P. Amato, A.H. van Kaam, D.G. Tingay, Z. Zhao, B. Grychtol, M. Bodenstein, H. Gagnon, S.H. Böhm, E. Teschner, O. Stenqvist, T. Mauri, V. Torsani, L. Camporota, A. Schibler, G.K. Wolf, D. Gommers, S. Leonhardt, A. Adler, T. study group, Chest electrical impedance tomography examination, data analysis, terminology, clinical use and recommendations: Consensus statement of the translational EIT development study group, in: E. Fan, W.R. Lionheart, T. Riedel, P.C. Rimensberger, F. Suarez Sipmann, N. Weiler, H. Wrigge (Eds.), Thorax 72 (1) (2017) 83–93.
[18] Y. Shi, Z. Yang, F. Xie, S. Ren, S. Xu, The research progress of electrical impedance tomography for lung monitoring, Front. Bioeng. Biotechnol. 9 (2021).
[19] B. Brazey, Y. Haddab, N. Zemiti, Robust imaging using electrical impedance tomography: Review of current tools, Proc. R. Soc. A Math. Phys. Eng. Sci. 478 (2258) (2022) 20210713.
[20] K. Haris, B. Vogt, C. Strodthoff, D. Pessoa, G.-A. Cheimariotis, B. Rocha, G. Petmezas, N. Weiler, R.P. Paiva, P. de Carvalho, N. Maglaveras, I. Frerichs, Identification and analysis of stable breathing periods in electrical impedance tomography recordings, Physiol. Meas. 42 (6) (2021) 64003.
[21] F.S. Hsu, S.R. Huang, C.W. Huang, C.J. Huang, Y.R. Cheng, C.C. Chen, J. Hsiao, C.W. Chen, L.C. Chen, Y.C. Lai, B.F. Hsu, N.J. Lin, W.L. Tsai, Y.L. Wu, T.L. Tseng, C.T. Tseng, Y.T. Chen, F. Lai, Benchmarking of eight recurrent neural network variants for breath phase and adventitious sound detection on a selfdeveloped open-access lung sound database-HF_Lung_V1, PLoS One 16 (7 July) (2021) 1–26.
[22] Yee Leng Yap, Z. Moussavi, Acoustic airflow estimation from tracheal sound power, in: IEEE CCECE2002. Canadian Conference on Electrical and Computer Engineering. Conference Proceedings, Vol. 2, Cat. No.02CH37373, IEEE, 2002, pp. 1073–1076.
[23] A. Yadollahi, Z. Moussavi, A robust method for estimating respiratory flow using tracheal sounds entropy, IEEE Trans. Biomed. Eng. 53 (4) (2006) 662–668.
[24] A. Yadollahi, Z.M.K. Moussavi, The effect of anthropometric variations on acoustical flow estimation: Proposing a novel approach for flow estimation without the need for individual calibration, IEEE Trans. Biomed. Eng. 58 (6) (2011) 1663–1670.
[25] L.L. Gomes, N.V. Oliveira, L.M. Tauil, R.A. Mattos, P.L. Melo, Instrumentation for respiratory flow estimation using tracheal sounds analysis: Design and evaluation in measurements of respiratory cycle periods and airflow amplitude, J. Phys. Conf. Ser. 1044 (1) (2018) 012037.
[26] D. Dellweg, P. Haidl, K. Siemon, P. Appelhans, D. Kohler, Impact of breathing pattern on work of breathing in healthy subjects and patients with COPD, Respiratory Physiol. Neurobiol. 161 (2) (2008) 197–200.
[27] C. Jácome, A. Marques, Computerized respiratory sounds: Novel outcomes for pulmonary rehabilitation in COPD, Respir. Care 62 (2) (2017) 199–208.
[28] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems, 2015, Software available from tensorflow.org.
[29] I. Frerichs, R. Paradiso, V. Kilintzis, B.M. Rocha, F. Braun, M. Rapin, L. Caldani, N. Beredimas, R. Trechlis, S. Suursalu, C. Strodthoff, D. Pessoa, O. Chételat, R.P. Paiva, P. de Carvalho, N. Maglaveras, N. Weiler, J. Wacker, Wearable pulmonary monitoring system with integrated functional lung imaging and chest sound recording: A clinical investigation in healthy subjects, Physiol. Meas. 44 (4) (2023) 045002.

[30] D. Pessoa, B.M. Rocha, C. Strodthoff, M. Gomes, G. Rodrigues, G. Petmezas, G.-A. Cheimariotis, V. Kilintzis, E. Kaimakamis, N. Maglaveras, A. Marques, I. Frerichs, P. de Carvalho, R.P. Paiva, BRACETS: Bimodal repository of auscultation coupled with electrical impedance thoracic signals, Comput. Methods Programs Biomed. 240 (2023) 107720.

[31] L.M.T. Jesus, A. Barney, R. Santos, J. Caetano, J. Jorge, P.S. Couto, Universidade de Aveiro's voice evaluation protocol, in: Interspeech 2009, May 2014, ISCA, ISCA, 2009, pp. 971–974.

[32] B. Brown, Electrical impedance tomography (EIT): A review, J. Med. Eng. Technol. 27 (3) (2003) 97–108.

[33] A. Adler, J.H. Arnold, R. Bayford, A. Borsic, B. Brown, P. Dixon, T.J. Faes, I. Frerichs, H. Gagnon, Y. Gärber, B. Grychtol, G. Hahn, W.R. Lionheart, A. Malik, R.P. Patterson, J. Stocks, A. Tizzard, N. Weiler, G.K. Wolf, GREIT: A unified approach to 2D linear EIT reconstruction of lung images, Physiol. Meas. 30 (6) (2009).

[34] D. Pessoa, B.M. Rocha, G.-A. Cheimariotis, K. Haris, C. Strodthoff, E. Kaimakamis, N. Maglaveras, I. Frerichs, P. de Carvalho, R.P. Paiva, Classification of electrical impedance tomography data using machine learning, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, 2021, pp. 349–353.

[35] A. Adler, W.R. Lionheart, Uses and abuses of EIDORS: An extensible software base for EIT, Physiol. Meas. 27 (5) (2006).

[36] T. Quatieri, Discrete-time speech signal processing: principles and practice, Pearson Education, 2002.

[37] B.M. Rocha, D. Pessoa, A. Marques, P. Carvalho, R.P. Paiva, Automatic classification of adventitious respiratory sounds: A (Un)solved problem? Sensors 21 (1) (2020) 57.

[38] C. Chourpiliadis, A. Bhardwaj, Physiology, respiratory rate — ncbi.nlm.nih.gov, 2022, https://www.ncbi.nlm.nih.gov/books/NBK537306/. (Accessed 02 Feb 2023).

[39] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, 2015, arXiv:1512.03385.

[40] H. Lim, J. Park, K. Lee, Y. Han, Rare sound event detection using 1D convolutional recurrent neural networks, in: Dcase 2017 Proceedings, November, 2017, pp. 2–6.

[41] H. Sakoe, S. Chiba, Dynamic programming algorithm optimization for spoken word recognition, IEEE Trans. Acoust. 26 (1) (1978) 43–49.

[42] B.M. Rocha, D. Filos, L. Mendes, I. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jácome, A. Marques, R.P. Paiva, I. Chouvarda, P. Carvalho, N. Maglaveras, A respiratory sound database for the development of automated classification, IFMBE Proc. 66 (2018) 33–37.

[43] B.M. Rocha, D. Filos, L. Mendes, G. Serbes, S. Ulukaya, Y.P. Kahya, N. Jakovljevic, T.L. Turukalo, I.M. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jácome, A. Marques, N. Maglaveras, R. Pedro Paiva, I. Chouvarda, P. de Carvalho, An open access database for the evaluation of respiratory sound classification algorithms, Physiol. Meas. 40 (3) (2019) 035001.

[44] L. Lasarow, B. Vogt, Z. Zhao, L. Balke, N. Weiler, I. Frerichs, Regional lung function measures determined by electrical impedance tomography during repetitive ventilation manoeuvres in patients with COPD, Physiol. Meas. 42 (1) (2021) 015008.

[45] B. Vogt, S. Pulletz, G. Elke, Z. Zhao, P. Zabel, N. Weiler, I. Frerichs, Spatial and temporal heterogeneity of regional lung ventilation determined by electrical impedance tomography during pulmonary function testing, J. Appl. Physiol. 113 (7) (2012) 1154–1161, PMID: 22898553.

[46] I. Frerichs, L. Lasarow, C. Strodthoff, B. Vogt, Z. Zhao, N. Weiler, Spatial ventilation inhomogeneity determined by electrical impedance tomography in patients with chronic obstructive lung disease, Front. Physiol. 12 (2021).

[47] B.M. Rocha, D. Pessoa, A. Marques, P. de Carvalho, R.P. Paiva, Automatic wheeze segmentation using harmonic-percussive source separation and empirical mode decomposition, IEEE J. Biomed. Health Inf. 27 (4) (2023) 1926–1934.