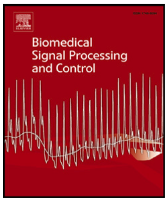
Biomedical Signal Processing and Control 92 (2024) 106009

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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



Pre-processing techniques to enhance the classification of lung sounds based on deep learning

Alessandra Fava [a], Behnood Dianat [a,b], Alessandro Bertacchini [a], Andreina Manfredi [d], Marco Sebastiani [c,d], Marco Modena [a], Fabrizio Pancaldi [a,b,*]

[a] Department of Sciences and Methods for Engineering, University of Modena and Reggio Emilia, via G. Amendola 2, 42122, Reggio Emilia, Italy
[b] Artificial Intelligence Research and Innovation Center (AIRI), University of Modena and Reggio Emilia, via P. Vivarelli 10, 41125, Modena, Italy
[c] Department of Surgery, Medicine, Dentistry and Morphological Sciences with Transplant Surgery, Oncology and Regenerative Medicine Relevance, University of Modena and Reggio Emilia, via del Pozzo 71, 41124, Modena, Italy
[d] Rheumatology Unit, Azienda Policlinico di Modena, via del Pozzo 71, 41124, Modena, Italy

ARTICLE INFO

Keywords:
Lung sounds
Deep learning

ABSTRACT

Deep learning has recently proved a huge potential in the classification of lung sounds. Most studies rely on publicly available data sets that are usually well-cleaned and annotated by expert physicians. The result of annotation is subjective by definition and, above all, large and public data sets are not collected in the scope of a very specific clinical investigation. Other works rely on private and suitably collected data sets that either may or may not stem from clinical studies. The main issue in these cases is represented by the reliability and noisiness of auscultations.

This paper delves into the significant impact of quantitative, systematic and reproducible cleaning of data sets of lung sounds. For “cleaning a data set” we mean discarding the records that carry mostly noise and interfering signals, since machine learning can be significantly impaired by outliers.

The developed pre-processing techniques are tested on several data sets of lung sounds. We designed a deep neural network (DNN) for the diagnosis of interstitial lung diseases (ILD) in patients affected by connective tissue diseases (CTD). The devised DNN can provide significant performance on the clean data set with impressive accuracy, F1-score, and F2-score of 97% with respect to the high-resolution computer tomography. Considering that the screening of ILD in patients affected by chronic autoimmune diseases is still an open issue, the proposed pipeline represents the enabling technology for the early, safe, reliable and cheap diagnosis of CTD-ILD.

1. Introduction

Interstitial lung disease (ILD) is one of the most severe and frequent complications of chronic autoimmune diseases like rheumatoid arthritis (RA) and connective tissue diseases (CTD). ILD appreciably reduces the life expectation of patients as well as it can worsen their quality of life. Despite these certainties, the actual incidence, prevalence and survival rate related to ILD are yet largely unknown and are mainly based on retrospective studies. Diagnosis of lung involvement in RA and CTD patients can be difficult if based only on symptoms. Patients can be asymptomatic in the early stages of the disease and for a long time, whereas some suggestive clinical manifestations, such as fatigue, dyspnoea and cough, can also derive from extra-pulmonary causes. High-resolution computed tomography (HRCT) remains the gold standard for the diagnosis of ILD and it is mandatory in case of suspected ILD. Nevertheless, a routine use of HRCT for screening programs is not advisable for both the high costs to be sustained by the national health system (NHS) and the exposition to ionizing radiation of patients. To improve the prescriptive appropriateness of HRCT for the early diagnosis of ILD, a physical lung examination has been proposed as an easy and repeatable screening. In fact, lung auscultation can reveal fine bibasilar, end-inspiratory, “velcro-like” crackles, which may precede the development of clinically overt ILD. Recently, our group developed an algorithm named VECTOR (VElcro Crackles detecTOR) capable of recognizing velcro crackles in pulmonary sounds with high sensitivity

* Corresponding author at: Department of Sciences and Methods for Engineering, University of Modena and Reggio Emilia, via G. Amendola 2, 42122, Reggio Emilia, Italy.

E-mail addresses: a.fava@unimore.it (A. Fava), behnood.dianat@unimore.it (B. Dianat), alessandro.bertacchini@unimore.it (A. Bertacchini), a.manfredi@unimore.it (A. Manfredi), marco.sebastiani@unimore.it (M. Sebastiani), marco.modena@unimore.it (M. Modena), fabrizio.pancaldi@unimore.it (F. Pancaldi).

https://doi.org/10.1016/j.bspc.2024.106009
Received 24 August 2023; Received in revised form 23 December 2023; Accepted 29 January 2024
Available online 7 February 2024

and specificity in both RA and CTD patients [1–3]. These algorithms are employed in a prospective study led by our research group and focused on the investigation of the incidence and prevalence of ILD in patients affected by RA and Sjogren's syndrome [4,5]. In practice, respiratory sounds are periodically recorded in 3 or 4 pulmonary fields (2 at the basal field, 1 at the middle field and eventually 1 at the upper field) in a silent environment with an electronic stethoscope. Considering a bilateral auscultation, 6 or 8 audio files are acquired for each patient.

The algorithmic classification of lung sounds has attracted much interest in the last 10 years. On the one hand, developing countries can take advantage of telemedicine to make up for the lack of specialized doctors in needy regions. On the other hand, physicians working in advanced hospitals or clinics can rely on quantitative tools to support their diagnosis. The huge potential of deep learning (DL) is giving a new life to this field of research. The pre-processing proposed in the work [6] consists of removing high-frequency components and zero-padding each auscultation to achieve a time support of 15 s. Then 13 features are extracted from the Mel-Frequency Cepstral Coefficients (MFCC) and 1000 features are extracted from the Short-Time Fourier Transform (STFT). The classifier is based on an Artificial Neural Network (ANN) achieving an accuracy of 98.61%. Empirical Mode Decomposition (EMD) and Gammatone filter bank are used in [7] for pre-processing. The classification relies on well-known architecture for neural networks, namely AlexNet, GoogLenet, ResNet50 and InceptionV3, where the respective accuracies are 98.60%, 98.80%, 98.80%, and 98.14%. MFCC are exploited for feature extraction and ANN is used for classification in [8]. Accuracies larger than 90% are obtained at high signal-to-noise ratios (SNRs). 3D-second order difference plot is investigated in [9,10] for feature extraction, whereas a deep autoencoder is considered for the classification of severity in patients affected by chronic obstructive pulmonary disease (COPD).

Most works rely on publicly available data sets that are usually well-cleaned and annotated by expert physicians. The workload required to specialized doctors is huge and the result of annotation is subjective by definition. Above all, large and public data sets are not collected in the scope of a very specific clinical investigation. For instance, the ICBHI 2017 challenge data set [11] described in [12] includes diagnosis of COPD, asthma, pneumonia and bronchiectasis. It is worth mentioning a couple of works processing the ICBHI 2017 challenge data set. In [13], MFCC and autoencoder are exploited for denoising, whereas classification is based on long-short term memory (LSTM) and bidirectional LSTM (BLSTM) networks. The accuracy is 94% for LSTM and 97% for BLSTM. In [14] several features are extracted from the data set, namely Shannon entropy, logarithmic energy entropy, and spectrogram-based spectral entropy. Decision tree (DT) and support vector machine (SVM) are employed for classification with an accuracy of 98% and 98.2%, respectively. Other works rely on private and suitably collected data sets that either may or may not stem from clinical studies. To the best of our experience, the main issue in these cases is represented by the reliability and noisiness of auscultations. In fact, patients affected by severe pulmonary disorders can only perform a few deep breaths before getting tired, so the skills of the medical staff play a fundamental role. Then, researchers and physicians have two options. The first option consists of considering the whole data set as is, however, "bad" auscultations will affect the final results. The second option consists of cleaning the data set on the basis of physicians' annotations, but the reproducibility of results is unavoidably compromised by the subjective judgment of specialized doctors.

On the one hand, the problem of noise suppression or noise reduction or noise mitigation in lung sounds has been widely investigated in the technical literature. For instance, the separation of heart sounds from lung sounds is a well known issue in clinical practice (see [15] and the references therein). The basic idea is that the useful signal is always available, but it might be hidden by noise; then, the goal is to filter out the noise component. The conventional approaches relies on adaptive filtering [15–17] and/or time–frequency analysis [15,18–22]. Active noise cancellation based on two microphones is employed in [23]. Mode decomposition has attracted much interest recently because of its ability to separate lung sounds from background noise [24–26]. On the other hand, the problem of cleaning data sets of lung sounds is still an under-explored field. To the best of our experience, in many clinical cases the useful signal is not available at all in every auscultation. For instance, when the pulmonary disease is very severe, the patient is not able to deeply breathe.

1.1. Scope of this work

To the best of our knowledge, the problem of quantitative, systematic and reproducible cleaning of data sets of lung sounds has not been tackled in the technical literature yet. For "cleaning a data set" we mean discarding the records that carry mostly noise and interfering signals, since machine learning can be significantly impaired by outliers. In this paper, we present an algorithmic approach to this problem. Variational mode decomposition (VMD) [7,27] and Harmonic Percussive Separation Spectrogram (HPSS) [28] are employed for signal denoising. Several features are extracted from the fast Fourier transform (FFT) and autocorrelation function of the resulting signal. Conventional techniques, namely K nearest neighbors (Knn) [29], Decision Tree (DT) [30], LogitBoost [31] and Naive Bayes (NB) [32], are used to classify auscultations into "good" and "bad" signals. We mean for good signal an auscultation carrying non-negligible information for the diagnosis of pulmonary disorders. Conversely, we categorize an auscultation as a 'bad signal' when it predominantly contains noise, lacking substantial diagnostic information. We consider two data sets collected in our previous studies and the publicly available RespiratoryDatabase@TR [33]. We compare the results of the proposed pipeline with a mixture of annotations provided by expert physicians and quantitative data devised from [3]. Then, bad auscultations are purged from the data sets and the performance of a new deep neural network (DNN) is assessed with respect to HRCT reports in the diagnosis of ILD.

1.2. Related works

A first example of work close to this topic is [34], where the classification of different noise sources and clean signals is considered in a population of young children. Several time–frequency features are extracted and classified through a SVM. A second example is the work [35] dealing with the discrimination between uncontaminated and noisy lung sounds. The analysis of [34] relies on spectrum characterization, whereas the analysis of [35] is based on Katz fractal dimension, Teager–Kaiser energy operator and normalized mutual information. However, excluding all the auscultations that carry some noise is not feasible in most cases, since this would lead to a significant reduction of the data set dimension and hence it would appreciably affect the performance of deep learning.

The remainder of the paper is organized as follows. The data sets considered for performance assessment are introduced in Section 2. The proposed pipeline for the classification of good and bad auscultations is described in Section 3. The new DNN for the detection of ILD is presented in Section 4. Numerical and experimental results for both the classification of good/bad sounds and the diagnosis of ILD are illustrated in Section 5. Finally, some conclusions are discussed in Section 6.

2. Clinical study

The data sets used in this work have been collected in the clinical studies described in detail in [4,5,36,37]. These studies were approved by the University Hospital of Modena Ethics Committee and all participants signed a written consent form. The studies were conducted in accordance with the principles of the Helsinki Declaration. In this Section, we recall only the information necessary to understand the composition of these data sets and to appreciate the practical impact of the results discussed in Section 5.

The population included in the RA-ILD data set involves patients affected by RA and undergoing chest HRCT at the University Hospital of Modena (Italy) in the period 1st of January 2015–1st of January 2017 [4]. HRCTs have been requested by physicians independent of this research and the patients have not been suitably selected. Lung auscultation was performed on both sides of the chest in three different areas (lower para-vertebral, lower axillary, medium para-vertebral) during an outpatient visit in a quiet environment. Littmann 3200 digital stethoscope was used to capture, convert, and record pulmonary sounds. The recordings were made at a sampling frequency of 4000 Hz. All 6 recordings per patient were saved as WAV (Waveform Audio files). The population included in the CTD-ILD data set involves patients with a diagnosis of CTD, such as dermatomyositis, Sjögren's syndrome, antisynthetase syndrome, systemic lupus erythematosus, and undifferentiated connective tissue disease [36]. Participants were required to have undergone an HRCT scan in the past 12 months before the study. Those with pleural effusion or pneumothorax were not eligible. In some cases, additional HRCT scans were performed. Lung auscultation was performed on both sides of the chest in four different areas (lower para-vertebral, lower axillary, medium para-vertebral and upper para-vertebral) in the same conditions and with the same setup of the RA-ILD study.

The RA-ILD data set is composed of 137 patients affected by RA and includes 820 auscultations. Similarly, the CTD-ILD data set is composed by 84 patients affected by CTD and includes 670 auscultations. Please see also Section 5 for the distribution of the data sets.

The HRCT images were anonymized, converted to DICOM format and analyzed by the same radiologist with specific expertise in thoracic radiology and ILD. The radiologist was required to extract binary information from HRCT images, namely the presence (positive) or absence (negative) of ILD. Radiological reports have been used as ground truth for the diagnosis of ILD in the performance assessment of the DNN devised in Section 4.

3. Classification of "good vs. bad" auscultations

In this Section, the suite of algorithms developed for the classification of "good vs. bad" auscultations is presented in detail. The designed pipeline is sketched in Fig. 1.

Variational mode decomposition (VMD) is used for separating different sources of sounds. We are interested in keeping lung sounds and removing all the other sounds, both physiological like heartbeat and artifacts like rubbing the stethoscope on the skin. The power of signals is normalized before VMD. Harmonic percussive separation spectrogram (HPSS) is introduced to remove percussive components of noise like tapping fingers on the head of the tool. The remaining harmonic component is divided into windows and the Root Mean Square (RMS) value is computed for each of them. The resulting sequence of RMS values (or RMS signal in brief) is interpolated through the Akima function [38]. Various features are extracted from the FFT and correlation function of the interpolated RMS signal, namely overall signal power, number of breath cycles, fundamental breathing frequency and breathing periodicity. Binary classifiers are employed to infer the presence of specific patterns related to breath cycles. Good auscultations include clean breath cycles and carry useful information for detecting ILD in patients affected by autoimmune diseases. On the contrary, bad auscultations mainly include noise and should not be used to raise the diagnostic suspicion of ILD.

The first step, i.e. VMD, is executed in Matlab environment [27], whereas the following steps are implemented in Python on the basis of the standard libraries Numpy, Pandas, Librosa, Path and Scipy [39–41].

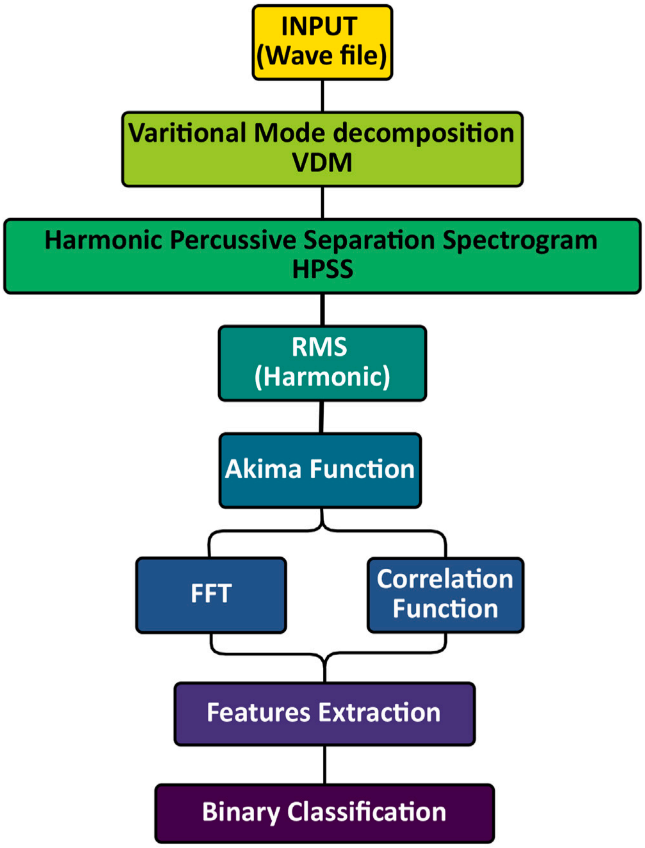


Fig. 1. Flow chart summarizing the pipeline for the detection of "good vs. bad" auscultations.

3.1. Variational mode decomposition

The first step of the pipeline consist of removing as much noise as possible. We usually classify the noise affecting auscultations as external and internal noise. External noise is mainly related to signal acquisition, for instance to the proper placement of the stethoscope head on the human body, to the pressure and tilt applied by the physician on the tool and to the electrical characteristics of the stethoscope. Internal noise is related to involuntary and voluntary physiological functions. To the best of our experience, involuntary physiological functions like heartbeat, stomach and intestine movements, are characterized by spectral components having frequencies lower than 100 Hz. Voluntary (in a wide sense) physiological functions like cough, crackles, wheezes and whistles, have spectral components with frequencies higher than 200 Hz.

In this work, we adopt VMD to separate different sources of lung sounds. Although Empirical Mode Decomposition (EMD) has been used in various applications [42–44], we have opted for VMD for its solid mathematical background. VMD [27] is a well-known approach to the non-recursive decomposition of a real input signal $g(t)$ into a discrete number K of sub-signals u_k. Sub-signals are called modes or Intrinsic Mode Functions (IMFs). The IMFs are amplitude-modulated–frequency-modulated (AM–FM) signals defined in [27] as

$$u_k(t) = A_k(t) \cdot cos(\phi_k(t)) \tag{1}$$

where the phase $\phi_k(t)$ is a non-decreasing function, the envelope $A_k(t)$ is non-negative, both the envelope and the instantaneous pulsation $\omega_k(t) = \phi'_k(t)$ vary much slower than the phase. The main goal consists of finding the modes $u_k(t)$ and respective central pulsation $\omega_k(t)$ to

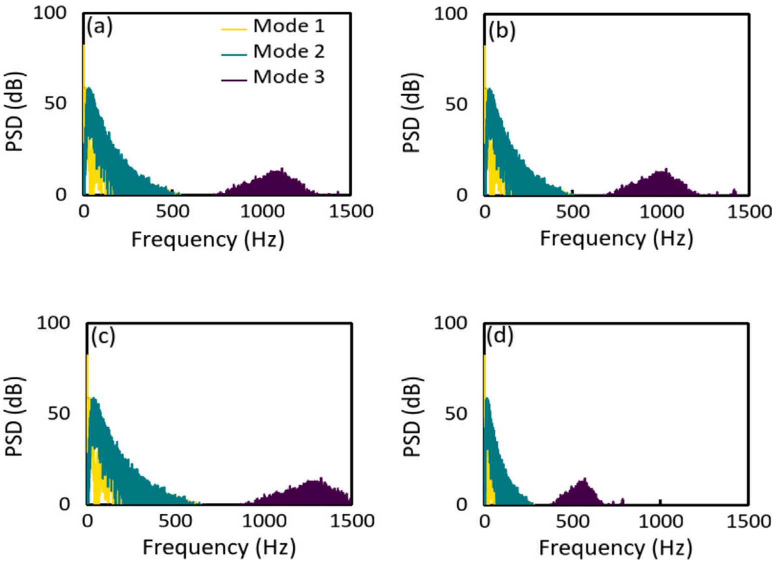


Fig. 2. PSD of the three IMFs resulting from the VMD of four auscultations: (a) good-positive, (b) bad-positive, (c) good-negative, (d) bad-negative.

minimize the variational problem

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad \text{s.t.} \quad \sum_k u_k = f \quad (2)$$

where $\delta(t)$ is the Dirac distribution, $*$ denotes convolution and $k = 1, 2, \ldots, K$. In practice, Hilbert Transform is used to devise the analytic signal corresponding to each mode, the mode is shifted to the base-band through the exponential of the given central frequency, the squared L2 norm of gradient is exploited to estimate the bandwidth of each mode [28]. The number of IMFs K to be devised represents an input for the algorithm and impacts the final result. On the one hand, if the number of expected IMFs is too low, a leakage between different sound sources may deteriorate the decomposition. On the other hand, an excessive number of IMFs may lead to noise over-fitting. In our work the number of IMFs has been set to $K = 3$ on the basis of both empirical tests on our data sets and the results of other studies in the same field. The IMFs at the output of the VMD are compact around the respective central pulsations $\omega_k(t)$. The parameters involved in VMD are described in detail in [27]. We used the Matlab function provided by D. Zosso for the practical implementation of VMD [45]. We set a moderate bandwidth constraint and the noise tolerance. The instantaneous frequencies are initialized as uniformly distributed.

According to the presented setup for VMD, the first mode mainly includes physiological functions and internal noise, so it is always discarded. The second mode usually includes a mixture of useful information and noise. If the central frequency of the second mode is lower than the frequency threshold $f_{th} = 110$ Hz, then it is discarded since it includes mainly noise. On the contrary, if the center frequency of the second mode is higher than the frequency threshold $f_{th} = 110$ Hz, it is considered since it carries useful information for this application. The frequency threshold $f_{th} = 110$ Hz has been set based on the doctor's annotations and empirical considerations [3]. The third mode usually carries the most useful information about physiological and pathological lung sounds. The third mode is discarded if its central frequency is lower than $f_{th} = 110$ Hz, since in this case, pulmonary sounds are almost absent. Indeed, the auscultation is classified as bad since it is meaningless to the scope of investigating pulmonary disorders. VMD evidenced an appreciable robustness against noise in some works available in the technical literature [7,46]. Summarizing, if the central frequency of mode 2 is larger than $f_{th} = 110$ Hz, the sum of the second and third IMFs is considered for further processing. If the central frequency of mode 3 is lower than $f_{th} = 110$ Hz, the auscultation is classified as bad. Otherwise, only the third mode is processed by the remaining of the pipeline.

The VMD of four auscultations is shown in Fig. 2. Left and right columns refer to good and bad auscultations, respectively. The top and bottom rows denote whether or not the patient is positive or negative to ILD, respectively. The power spectral density (PSD) of the three IMFs is expressed as dBW/Hz. The central frequency of the modes increases as the mode order increases by the definition of VMD. Furthermore, the power of the modes decreases as the mode order increases, since lung sounds are weaker than other physiological sounds and possible artifacts/noise. The overall "pictures" of the four VMDs are quite similar, even if some quantitative deviations can be appreciated in the bad-negative auscultation. Consequently, further processing is necessary for the classification of lung sounds.

3.2. Mel spectrogram

Mel spectrogram is exploited in this Section to visualize the time–frequency properties of good and bad auscultations. In our Python implementation, we focus on the first 8 s of the considered mode/modes (see Section 3.1) for two main reasons. Firstly, this time support is suitable to collect at least two to four breath cycles [3]. Secondly, inhalations are usually deeper at the beginning of the auscultation, i.e. when the patient is not "tired" yet. It is well known that the deeper is the inhalation, the more likely is the detection of possible velcro crackles since the pathogenesis of ILD starts in the lower lobes. Each mode/sum of modes is divided into 500 non-overlapping frames of 64 samples, considering the sampling rate of 4000 Hz for the electronic stethoscope employed in this study. The power of the considered signals has been normalized to compare the auscultations acquired by different operators in different hospitals. This approach has been exploited for cough recognition in [47] and for the detection of ILD in [3].

The Mel spectrograms [48] of the IMFs depicted in Fig. 2 are shown in Figs. 3 and 4. Auscultations referring to patients positive and negative to ILD are illustrated in Figs. 3 and 4, respectively. Good and bad auscultations are reported in the left and right columns, respectively, of both figures. Similarly, the first row of both figures represents the acquired signal, and the second to fourth rows denote first to third IMFs, respectively.

Good signals are characterized by almost periodic breath cycles in both the raw acquisition and IMFs (left columns of Figs. 3 and 4). Most of the power of the first and second IMFs is carried in the band 0−100 Hz, whereas the power of the third IMF is gathered around

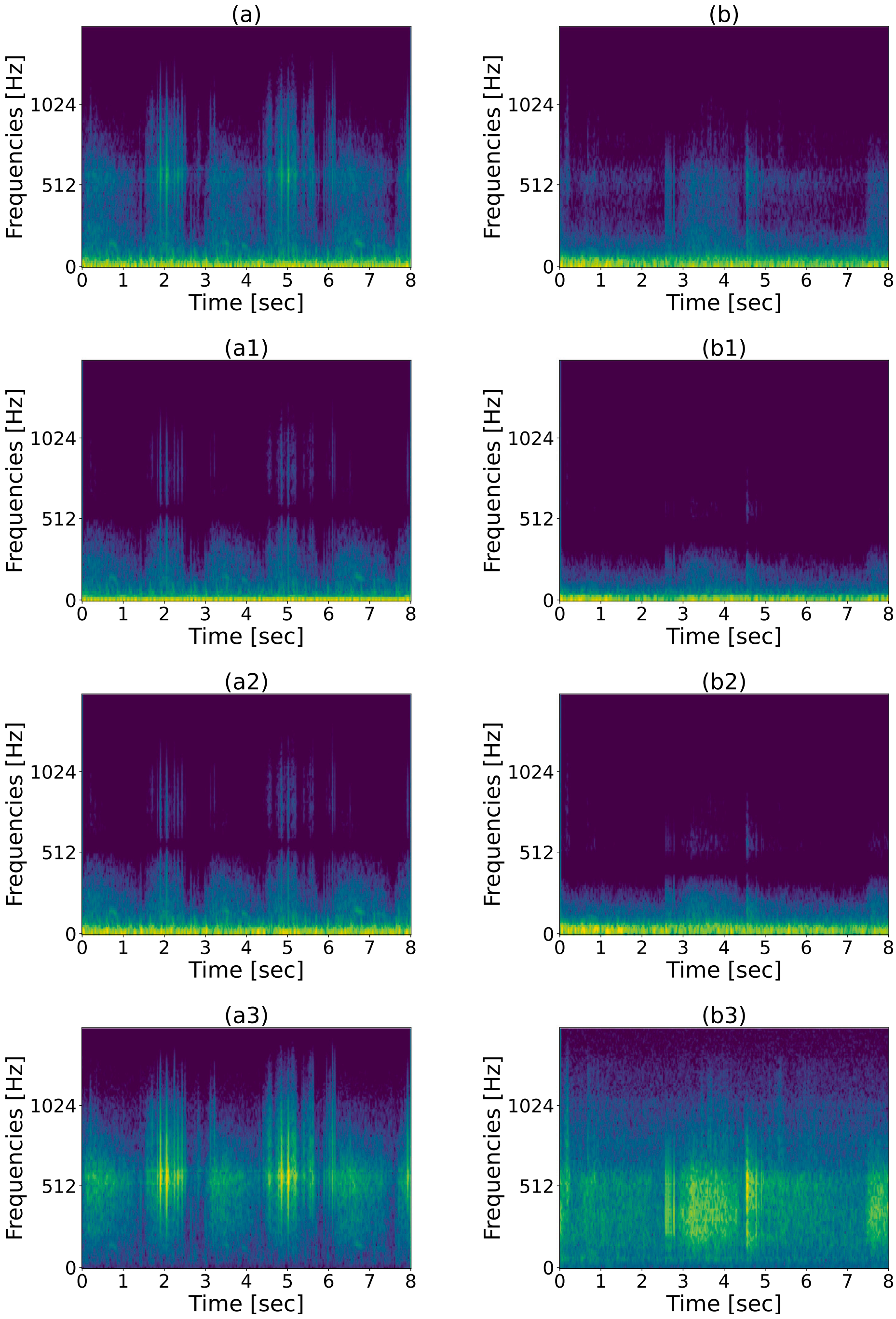


Fig. 3. Mel spectrograms of auscultations referring to patients positive to ILD (first row of Fig. 2): (a) full-signal good-positive, (b) full-signal bad-positive, (a1) mode 1 good-positive, (b1) mode 1 bad-positive, (a2) mode 2 good-positive, (b2) mode 2 bad-positive, (a3) mode 3 good-positive, (b3) mode 3 bad-positive.

500 Hz with strong components up to 1000 Hz. On the contrary, bad signals evidence irregular breathing patterns as shown in the right columns of Figs. 3 and 4. Most of the noise affects the first and second IMFs. The power of the third IMF is smeared in the time–frequency plane. Events characterized by very short time support and frequencies up to 1500 Hz can be easily related to artifacts, like for instance tipping the fingers on the head of the stethoscope.

3.3. Harmonic percussive separation spectrogram

HPSS [28] is employed for further noise suppression. HPSS is adopted in several fields to evidence the harmonic vs. percussive components in a spectrogram, for instance, [49]. Percussive components are basically strong events concentrated in a limited time support and appear in the spectrogram as vertical lines. Harmonic components infer the presence of given spectral contributions and appear in the spectrogram as horizontal lines [28]. In this work, we are interested in harmonic components, since they can be related to signal components having a periodic behavior. On the contrary, percussive components are neglected since they are related to impulsive events usually associated with artifacts and noise sources. Not all sounds evidence necessarily harmonics and/or percussive components. The traditional audio model is based on the superposition of sines, transients, and noise (STN), so it is also called the STN model [28,50,51]. In practice, by discarding the first and second IMFs, we aim to filter out transients and noise, while retaining the sine parts that carry crucial information about

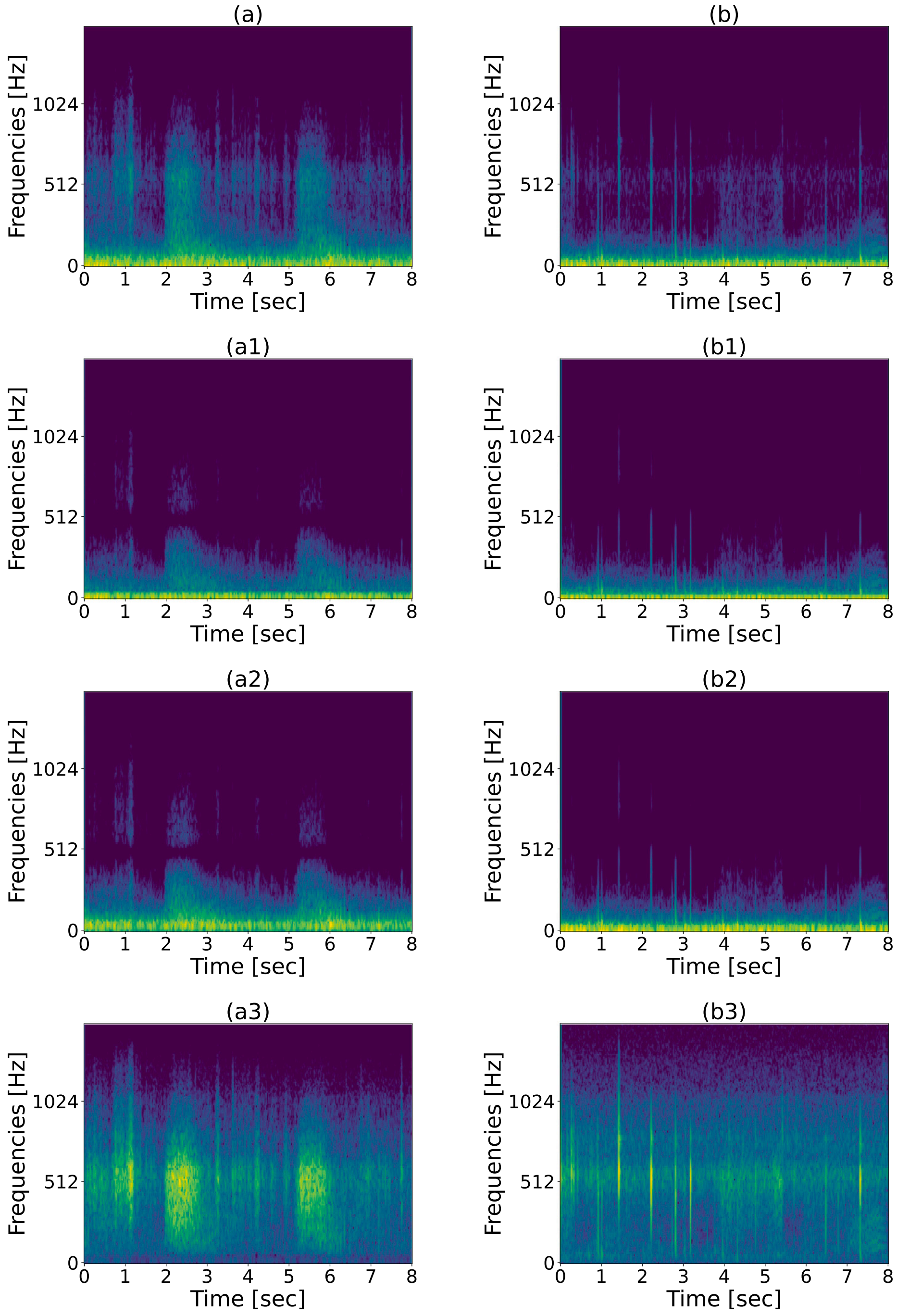


Fig. 4. Mel spectrograms of auscultations referring to patients negative to ILD (second row of Fig. 2): (a) full-signal good-negative, (b) full-signal bad-negative, (a1) mode 1 good-negative, (b1) mode 1 bad-negative, (a2) mode 2 good-negative, (b2) mode 2 bad-negative, (a3) mode 3 good-negative, (b3) mode 3 bad-negative.

pathological lung sounds. We employ HPSS to further evidencing harmonic components with respect to other components.

The mathematical details of HPSS are available in [28]. Denoting with

$$X(t,k) = \sum_{n=0}^{N-1} w(n)x(n + tH)exp(-2\pi \cdot ikn/N) \tag{3}$$

the STFT of either the third IMF or the mixture of the second and third IMFs (see Section 3.1), then the spectrograms of the harmonic and percussive components are defined as

$$X_h(t,k) = X(t,k) \cdot M_h(t,k) \tag{4}$$

and

$$X_p(t,k) = X(t,k) \cdot M_p(t,k), \tag{5}$$

respectively, where

$$M_h(t,k) = Y_h^1(t,k)/(Y_p^1(t,k) + \epsilon) > \beta \tag{6}$$

and

$$M_p(t,k) = Y_p^1(t,k)/(Y_h^1(t,k) + \epsilon) \geq \beta \tag{7}$$

are the masks for harmonic and percussive separation, respectively,

$$Y_h^1(t,k) = median(Y(t - \ lh,k),\dots,Y(t+lh,k)) \tag{8}$$

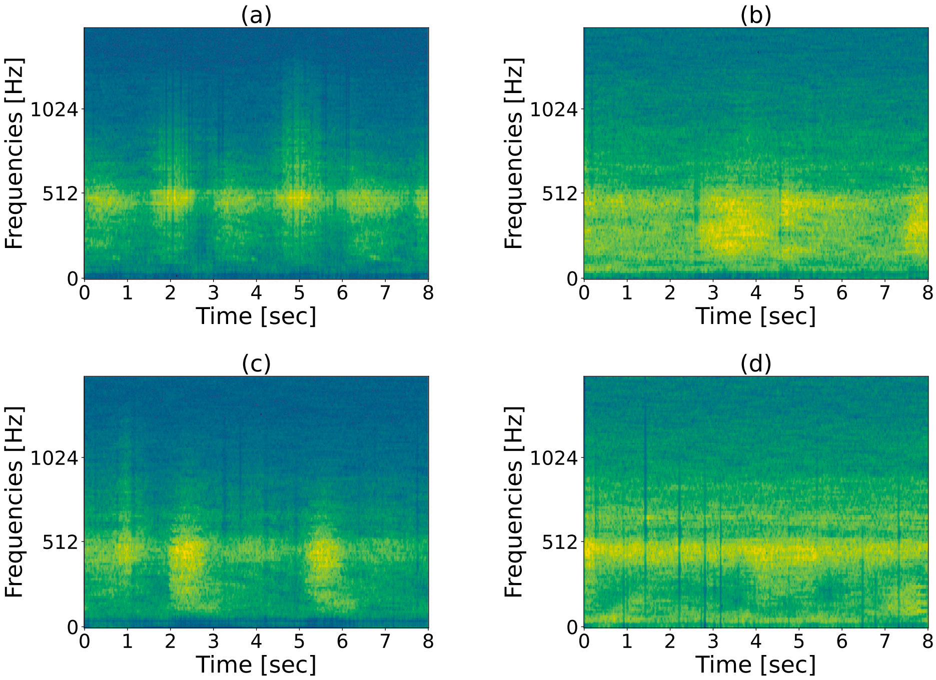


Fig. 5. Harmonic components of the third IMF of the signals considered in Figs. 2–4: (a) good-positive, (b) bad-positive, (c) good-negative, (d) bad-negative.

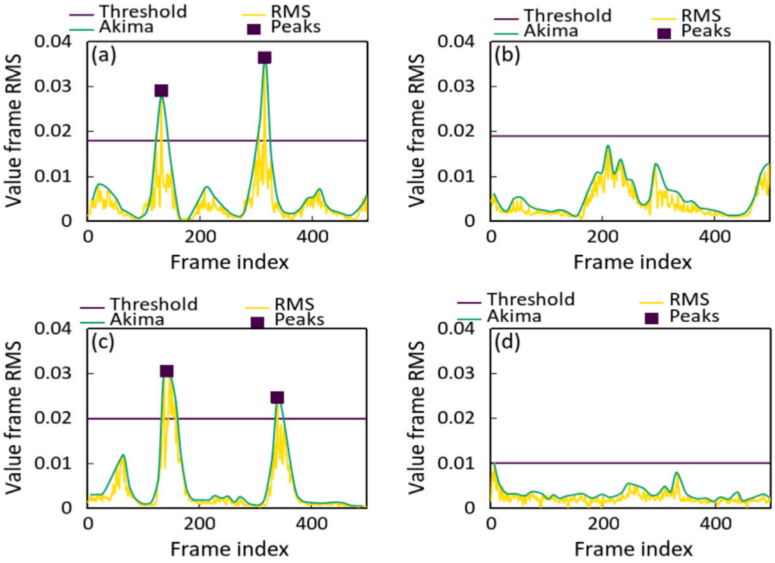


Fig. 6. RMS signals stemming from the harmonic components of Fig. 5: (a) good-positive, (b) bad-positive, (c) good-negative, (d) bad-negative. The yellow line represents the RMS signal, green line denotes Akima interpolation and blue circles highlight peaks.

and

$$Y_p^1(t,k) = median(Y(t,k-\ lp),\ldots,Y(t,k+lp)) \tag{9}$$

are the harmonically and percussively enhanced magnitude spectrograms, respectively, and $Y = |X|$. The separation factor of the masks has been set to $\beta = 2$ on the basis of both the indications of [28] and our empirical tests of robustness. The window dimension and the hop size have been set to $N = 256$ and $H = 192$ according to various studies available in technical literature [52–54]. Kaiser windowing has been employed for $w(n)$.

Fig. 5 presents the harmonic components of the third IMFs of the signals considered in Figs. 2, with good signals on the left and bad signals on the right. The good signals are characterized by numerous horizontal lines, which represent harmonic components. These components, particularly dense in time intervals corresponding to inhalations, facilitate the identification of breath cycles. Conversely, bad signals do not exhibit well-defined patterns, making breath cycles difficult to discern.

3.4. RMS

The root mean square (RMS) value of each time window composed by $N = 256$ samples is computed and then 1D Akima interpolation [38] is applied. This interpolation is employed to preserve crucial information about peaks, as the Akima function connects the peaks of the signal using splines. The output of the Akima interpolator is henceforth referred to as the RMS signal.

The RMS signals stemming from the harmonic components of Fig. 5 are shown in Fig. 6. Yellow line represents the RMS signal, green line denotes Akima interpolation and blue circles highlight peaks. Abscissa

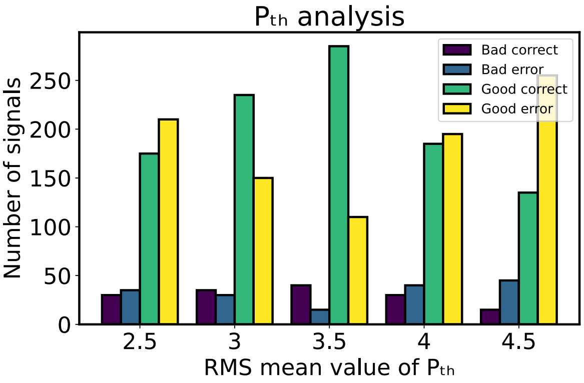


Fig. 7. Comparison between algorithmic classification and physicians' annotations ("good" vs. "bad" auscultations) for different power thresholds P_{th}.

and ordinate represent the frame index and RMS value (of the given frame), respectively. The amplitude of the peaks in good signals (on the left) is larger than that of the bad signals (on the right). Moreover, peaks in good signals are almost periodic in time and can be exploited to infer the presence of breathing activity.

3.5. Features extraction

The foundational premise of this work is the characterization of breathing as a pseudo-periodic process. The algorithms detailed in Sections 3.1 and 3.3 are specifically designed to eliminate transient and noise components from the conventional STN audio model while amplifying the sine (or harmonic) component. The subsequent sections will introduce the algorithms we propose for extracting various features of interest from both the raw signal (i.e., the acquired signal) and the RMS signal discussed in Section 3.4. In particular, the vector at the input of the classifier is composed by 6 features, namely signal power (Section 3.5.1), number of breath cycles (Section 3.5.2), frequency and amplitude of the absolute maximum of the FFT applied to the RMS signal (Section 3.5.3), amplitude and lag of the non-trivial maximum of the autocorrelation function (Section 3.5.4).

3.5.1. Power

This feature consists of the original power of the acquired signal and is used for power normalization.

3.5.2. Number of breath cycles

This feature quantifies the number of breath cycles. Specifically, the number of peaks in the Akima function that exceed a power threshold P_{th} are identified as breath cycles. The performance of the proposed algorithm for the detection of "good" vs. "bad" auscultations have been compared to the annotations of physicians for some values of the power threshold, namely for $P_{th} = \{2.5, 3, 3.5, 4, 4.5\} \cdot \overline{RMS}$, where $\overline{RMS}$ represents the mean value of the signal RMS. Results are summarized in Fig. 7 showing the number of correct vs. wrong decisions of the proposed algorithm for different power thresholds P_{th}. Correct decisions correspond to properly detecting good or bad auscultations, whereas wrong decisions lead to misdetection. It is worth pointing out that the search space might be expanded and enriched, however, this "brute force" approach requires significant efforts, so we limited our search to 5 values in the most promising set based on preliminary results. The best performance is then achieved setting the power threshold to

$$P_{th} = 3.5 \cdot \overline{RMS}. \tag{10}$$

Fig. 6 provides an example of peak counting. Two peaks are detected in the good signals illustrated in Figs. 6-(a) and -(c), whereas no peaks are identified in the bad signals depicted in Figs. 6-(b) and -(d).

3.5.3. FFT

Fast Fourier Transform (FFT) is exploited to devise the fundamental breathing frequency. Although the original breath signal is non-stationary and FFT is applied to the RMS signal, a certain pseudo-periodicity is still expected. In fact the auscultation length is 8 s and 2–4 breath cycles are expected (see Section 3.2). To this aim, the RMS signal of length 500 samples is included in one FFT window of length 2500 samples and zero-padding is employed to refine the spectral representation.

Useful information is carried by both the amplitude and frequency of the largest peak. Amplitude is interpreted as a measure of feature reliability. In fact, large amplitudes are achieved when the breathing is pseudo-periodic, whereas low amplitudes denote noise and/or absence of breathing cycles. In some cases, the FFT of the RMS signal evidences one peak corresponding to the breath fundamental frequency. In other cases, the FFT reveals two distinct peaks corresponding to different inhaling and exhaling pseudo-periodicity.

The FFT of the RMS signals considered in Fig. 6 is shown in Fig. 8. The FFTs associated with good signals (on the left) are characterized by pronounced peaks having normalized amplitude 1.5 and frequency 0.6 Hz. This frequency leads to about 5 breath cycles in the time interval of 8 s, which is a physiological condition. The FFTs associated to bad signals (on the right) are characterized by peaks weaker than those related to good signals. Moreover, the frequency of the main peaks are around 0.15 Hz, leading to about 1 breath cycle over a time interval of 8 s, i.e. evidencing an abnormal breathing condition.

3.5.4. Correlation function

The auto-correlation function is employed to analyze the pseudo-periodicity of the RMS signal (refer to Section 3.4). For instance, Fig. 9 illustrates the auto-correlation functions of the RMS signals considered in Fig. 6. Good auscultations, shown on the left, display a broad sense of pseudo-periodic behavior. This can be leveraged to infer physiological breathing. Indeed, the lag of the peaks is associated with the duration of breathing cycles, while the amplitude of the peaks can serve as a measure of estimation reliability. Conversely, bad auscultations, shown on the right, exhibit weak auto-correlation and a lack of periodicity. These characteristics can be used to infer the prevalence of noise over pulmonary sounds.

3.6. Classification

The features extracted from both the raw and RMS signals serve as inputs for various binary classification algorithms, which distinguish between "good" and "bad" auscultations. We have explored the use of several classifiers, including K-nearest neighbors (Knn) [29], Decision Tree (DT) [30], LogitBoost [31], and Naive Bayes (NB) [32]. These techniques find application in a multitude of fields, such as the classification of breast cancer metastasis [55].

3.7. Computational complexity

The computational complexity involved by the techniques presented in Section 3 has been assessed in terms of complex additions and multiplications. The pipeline has been sectioned into three main submodules, namely VMD, features extraction and classification. We denote with N_a the number of audio files, M_a is the average length of audio files, L_m is the average length of the modes, $P = 6$ is the number of features.

VMD has a computational complexity of $O(N_a \cdot M_a)$, whereas sorting and extraction of modes involve $O(N_a \cdot log(L_m))$ complex additions and multiplications.

The extraction of power, number of breath cycles and autocorrelation features require $O(N_a \cdot L_m)$ real operations, whereas FFT involves $O(L_m \cdot log(L_m))$ complex additions and multiplications.

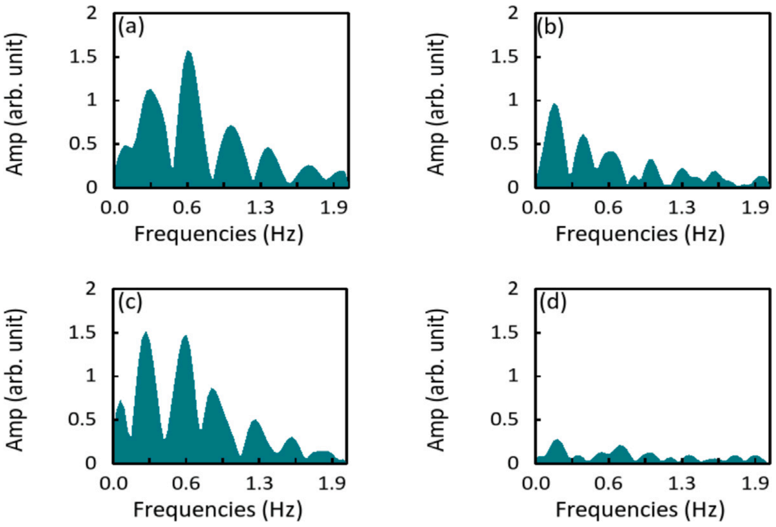


Fig. 8. FFT of the RMS signals of Fig. 6: (a) good-positive, (b) bad-positive, (c) good-negative, (d) bad-negative.

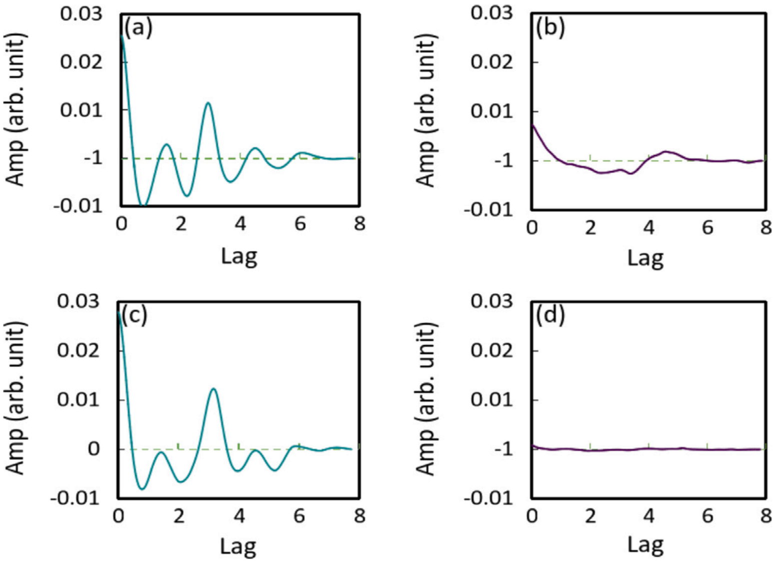


Fig. 9. Auto-correlation function of the RMS signals of Fig. 6: (a) good-positive, (b) bad-positive, (c) good-negative, (d) bad-negative.

Knn has a computational complexity of $O(N_a \cdot P \cdot K_n)$, where K_n is the number of neighbors. Logit Boost has a computational complexity of $O(N_a \cdot P \cdot Q)$, where Q represents the number of boosting iterations. Decision Tree involves a computational complexity of $O(P \cdot N_a \cdot log(N_a))$. Naive Bayes has a computational complexity of $O(N_a \cdot P \cdot K_b)$, where K_b is the number of classes.

4. Deep learning for the diagnosis of ILD

Our research group has extensively investigated the diagnosis of ILD from the analysis of lung sounds in several studies [1,3,4]. In particular, the work [3] introduced a pipeline based on deep learning that considers the whole available data set; in other words, suboptimal signals, i.e. bad auscultations, are not differentiated or removed. Despite this, in this work we prove that carefully curating and excluding unsuitable signals can significantly enhance the accuracy of ILD detection.

The core of our DNN consists of a multi-model parallel ensemble of well-known convolutional neural networks (CNNs) architectures, including InceptionV3, ResNet101, EfficientNetB0, and MobileNetV2. These CNNs are fed with an augmented image data set composed by the Mel spectrograms of lung sounds [3]. Each Mel spectrogram is handled as an image providing a standardized visual representation of the pulmonary sounds of the patient. The proposed DNN is trained exclusively on data unrelated to the test subjects to ensure an unbiased training process. In practice the training and testing data sets are strictly distinct.

Each CNN model is fine-tuned in its final layer for the specific task of ILD detection. Global average pooling is employed to extract key features; these are then processed through fully connected layers with dropout regularization to mitigate the risk of overfitting. The sigmoid activation function is adopted in the final layer to yield the class prediction. The individual predictions from the four CNN models are aggregated and further processed through additional fully connected layers with dropout. The resultant stacked architecture culminates in a sigmoid-activated layer producing the ensemble prediction.

This sophisticated model operates on the TensorFlow and Keras frameworks and is optimized using the Adam algorithm. Cross-entropy loss function is employed for training. 5-fold cross-validation is adopted for robust and reliable performance assessment in the detection of ILD from lung sounds.

Table 1
Performance of binary classifiers introduced in Section 3.6 for "good vs. bad" signals on the CTD-ILD data set. The accuracy is reported in terms of the mean and standard deviation of the cross-validation.

Metrics	Knn	LogitBoost	NB	DT
Accuracy	95% ± 0.45%	95% ± 0.84%	93% ± 0.55%	94% ± 0.74%
F1-score	97% ± 0.24%	97% ± 0.44%	96% ± 0.30%	97% ± 0.39%
F2-score	97% ± 0.34%	97% ± 0.60%	95% ± 0.44%	97% ± 0.50%
Recall	98% ± 0.42%	97% ± 0.74%	94% ± 0.53%	97% ± 0.60%
Precision	96% ± 0.17%	97% ± 0.40%	99% ± 0.12%	97% ± 0.35%

Table 2
Results of the Shapiro–Wilk test on the proposed classifiers.

Shapiro–Wilk test	Knn	LogitBoost	NB	DT
Statistic	0.9945	0.9939	0.9925	0.9618
p-value	0.9601	0.9383	0.8586	0.0054

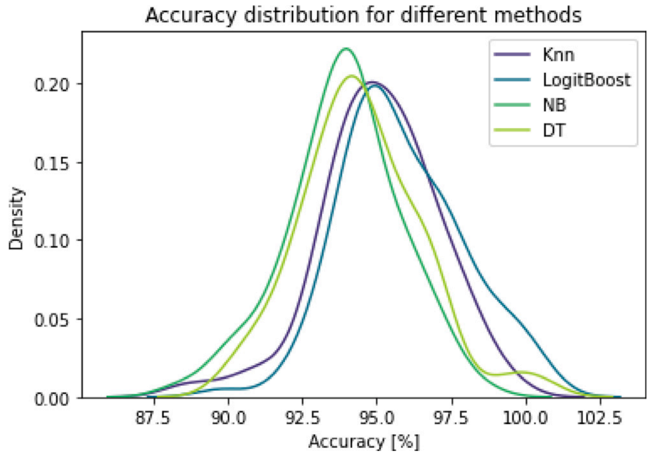


Fig. 10. Visual representation of the normal probability density functions corresponding to the Shapiro–Wilk test results for the classifiers Knn, LogitBoost, NB, and DT.

Table 3
Performance of binary classifiers introduced in Section 3.6 for "good vs. bad" signals on the RA-ILD data set. The maximum uncertainty entailed by the five-fold cross validation is ±5%.

Metrics	Knn	LogitBoost	NB	DT
Accuracy	58%	52%	58%	57%
F1-score	73%	73%	71%	73%
F2-score	86%	84%	80%	85%
Recall	97%	94%	87%	96%
Precision	59%	59%	60%	58%

5. Results

This Section is divided into three parts. First, the performance of the pipeline described in Section 3 is shown to "clean" the considered data sets of lung sounds (see Section 2), namely the CTD-ILD data set, the RA-ILD data set and the RespiratoryDatabase@TR. These results are presented in Sections 5.1 and 5.2. Then, the clean data sets for CTD-ILD and RA-ILD are exploited to raise the suspicion of ILD from lung sounds in Section 5.3. Finally, the results of Sections 5.1 and 5.3 are compared to the results presented in similar works in Section 5.4.

5.1. Cleaning of CTD-ILD and RA-ILD data sets

The process of classifying signals into "bad" and "good" categories can be deemed subjective, as it lacks a well-defined ground truth. To address this issue, we combine expert annotations from medical specialists and predictions from the DNN described in Section 4. Specifically, we define "bad" and "good" signals in the following way. We label an auscultation as "bad" if it lacks meaningful information according to medical specialists, and if the DNN prediction falls below 60%. Any auscultation not meeting these criteria is categorized as "good".

The CTD-ILD data set is unbalanced, with nearly twice as many 'good' signals compared to 'bad' ones. To counteract this issue, we assigned double the cost to 'bad' signals compared to 'good' ones, helping to reduce potential bias in the signal classifiers. We evaluated the performance using common metrics: accuracy, recall, precision, F1-score, and F2-score. Here, 'good' signals were treated as true positives and 'bad' signals as true negatives.

The performance of the pipeline described in Section 3 is summarized in Table 1 for the CTD-ILD data set. 5-fold cross-validation has been adopted for binary classification. The probability density functions (pdfs) entailed by the considered classifiers are shown in Fig. 10, whereas the results of the normality Shapiro–Wilk test are reported in Table 2. All the considered classifiers provide a normal accuracy distribution, except DT which can be deemed approximately normal. The performance is excellent for all classifiers, as all the metrics exceed 93%. In other words, all the classifiers are capable of identifying good/bad auscultations in the CTD-ILD data set with respect to the ground truth defined at the beginning of this Section. For the sake of fairness, Knn and LogitBoost provide the best performance exceeding 95% in all the metrics. Knn relies less than its counterparts on hyperparameters and is known for its simplicity, flexibility and accuracy. LogitBoost belongs to the category of ensemble methods, i.e. it combines different models to boost reliability, and it often yields excellent results. NB methods rely on the assumption that features are independent, but in our application features are somewhat correlated as discussed in Section 3.5. This correlation unavoidably affects the performance of the NB approach. DT method is more prone to overfitting than Knn and LogitBoost and suffers from a performance deterioration accordingly.

The same pipeline of Section 3 has been tested on the RA-ILD data set. The resulting performance is summarized in Table 3. Recall and F2-score exceed 94% and 80% respectively, denoting a minimal number of false negatives. In other words, the proposed pipeline is effective in identifying almost all good signals. However, precision is limited to 59-60%, denoting a significant number of false positives. Indeed, precision also affects accuracy and F1-score. In practice, several auscultations are classified as good even if they do not carry useful information for the diagnosis of ILD. We are aware that the RA-ILD data set collected in our first clinical study [1] may have some limitations. We suspect that our confidence in the electronic stethoscope and measurement setup was not optimal during data acquisition. Furthermore, the clinical picture of RA patients is usually more severe than that of CTD patients and repeatedly deep breathing is more difficult for them.

5.2. Cleaning of the RespiratoryDatabase@TR

The pipeline presented in Section 3 has been applied to the public data set available at [33] and described in [56]. Knn, Logit Boost and DT classifiers detected 3 "bad" auscultations over 504 files. Bad auscultations belongs to the same 4 patients. From a subjective and perceptive analysis of physicians, these bad auscultations have been probably discarded since breathing cycles cannot be detected. NB detected 14 "bad" auscultations over 504 files. From a subjective and perceptive analysis of physicians, these bad auscultations have been probably discarded for a combination of weak breath sound, cough and artifacts.

5.3. Diagnosis of ILD

The diagnosis of ILD is based on the HRCT as explained in Section 2. True positives denote patients affected by ILD, true negatives represent

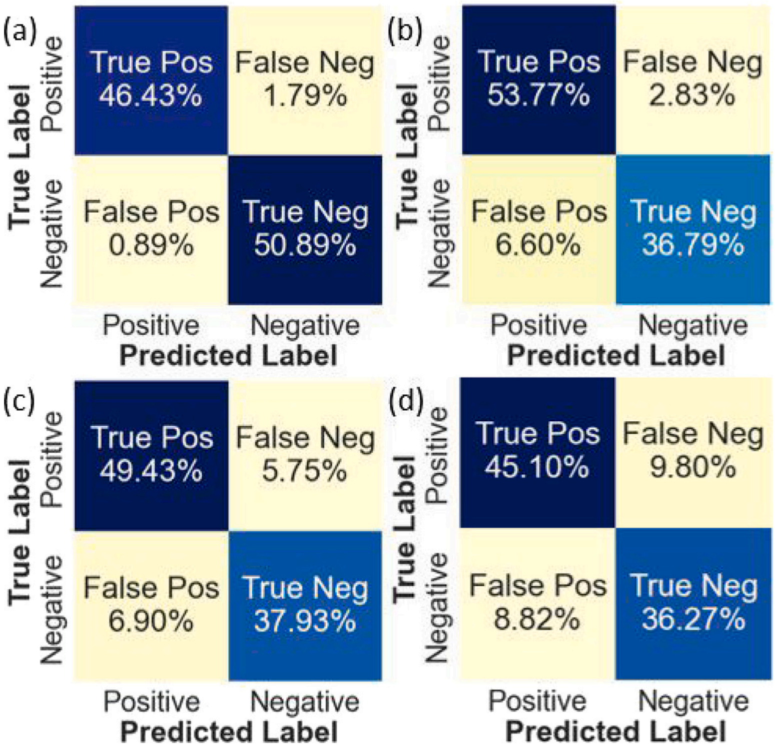


Fig. 11. Confusion matrices of the DNN processing the clean CTD-ILD data set: (a) Knn, (b) LogitBoost, (c) NB, (d) DT. True positives denote patients affected by ILD, true negatives represent patients not affected by ILD.

Table 4
Summary of the performance of the DNN in the diagnosis of ILD on the CTD-ILD data set. The maximum uncertainty entailed by the five-fold cross validation is ±2%.

Metric	Raw	Knn	LogitBoost	NB	DT
Accuracy	88%	97%	91%	87%	81%
F1-score	88%	97%	91%	87%	81%
F2-score	93%	97%	94%	89%	82%
Recall	94%	96%	95%	90%	82%
Precision	87%	98%	89%	88%	84%

patients not affected by ILD. Each data set, namely CTD-ILD and RA-ILD, goes through the pipeline of Section 3 yielding 4 distinct data sets, one for each classification method (see Section 3.6), namely Knn, LogiBoost, NB and DT. The raw, i.e. original, data set is considered for comparison. These distinct data sets feed the DNN presented in Section 4.

The confusion matrices related to the CTD-ILD data sets are shown in Fig. 11, whereas the performance metrics of the DNN are summarized in Table 4. The proposed pipeline for data cleaning can provide significant performance improvement in the diagnosis of ILD with respect to the raw, i.e. original, data set. The most significant improvement is entailed by the Knn classifier. This leads to a data set suitable to the DNN for achieving F1-score and F2-score of 97.4% and 96.7%, respectively, whereas the scores of the DNN on the original CTD-ILD data set are 88.1% and 92.8%. The LogitBoost classifier can provide a limited performance improvement, in fact, the related F1-score and F2-score are 90.5% and 93.8%, respectively. NB and DT cannot work well enough to increase the performance of the DNN in the diagnosis of ILD. Despite the similar capability of Knn and LogitBoost in classifying good and bad signals (see Table 1), their different performance in the diagnosis of ILD can be attributed to their inherent properties. Knn can effectively capture the underlying patterns and remove outliers by considering local similarity among neighboring samples. The resulting data set provides a more accurate representation of patterns, so that the DNN can be better trained. On the other hand, LogitBoost aims at optimizing the overall predictive performance and may not be adequate to effectively remove outliers. The remaining misclassified instances hinder the classification capability of the DNN.

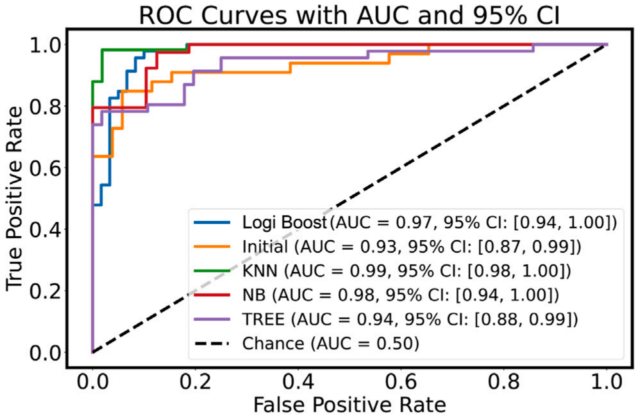


Fig. 12. ROC/AUC diagram of the considered data sets related to CTD-ILD. The 95% confidence interval (CI) is also shown to ease the comparison.

The results summarized in Table 4 are confirmed by the receiver operating characteristic/area under curve (ROC/AUC) diagram of Fig. 12. The 95% confidence interval (CI) is also shown to ease the comparison. Knn involves the largest AUC of 0.99. LogiBoost data set outperforms the raw CTD-ILD data set with an AUC of 0.97 versus 0.93, respectively. NB and DT deserve a particular discussion, since in these cases the DNN was impaired by the large amount of instances labeled as ‘bad’ reducing the size of the useful data set. The severe imbalance between ‘good’ and ‘bad’ instances in the raw data set poses further challenges. Although 5-fold validation was applied to mitigate classification uncertainty, large output fluctuations are observed and the DNN leads to inconsistent AUC/ROC values. As a consequence, despite the AUC of NB is as high

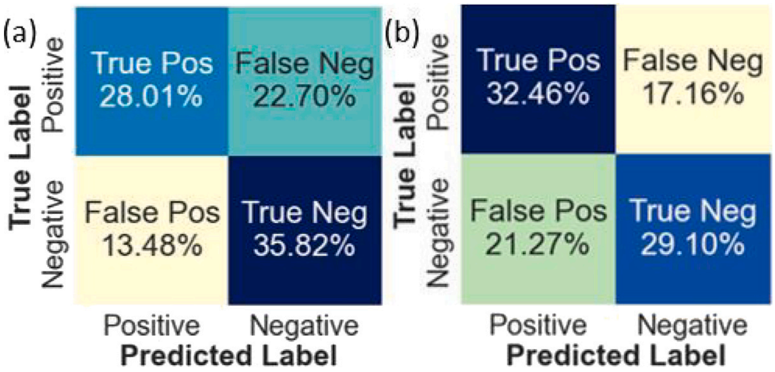


Fig. 13. Confusion matrices of the DNN processing the clean RA-ILD data set: (a) Knn, (b) LogitBoost. True positives denote patients affected by ILD, true negatives represent patients not affected by ILD.

Table 5
Summary of the performance of the DNN in the diagnosis of ILD on the RA-ILD data set.

Metrics	Only DNN	Knn	LogitBoost
Accuracy	68% ± 1%	64% ± 1%	62% ± 2%
F1-score	68% ± 2%	64% ± 1%	61% ± 1%
F2-score	62% ± 2%	57% ± 1%	64% ± 1%
Precision	65% ± 1%	68% ± 2%	60% ± 1%
Recall	61% ± 2%	55% ± 1%	65% ± 2%

as 0.98, the performance of the DNN on the NB data set is poor. The application of data balancing techniques might improve the consistency and performance of NB and DT in the context of ILD diagnosis, at least partially.

The performance of the developed DNN has been also assessed with respect to the RA-ILD data set. The confusion matrices are shown in Fig. 13, whereas the DNN performance metrics are summarized in Table 5. Only Knn and LogitBoost classifiers are considered for cleaning the RA-ILD data set, as NB and DT cannot provide satisfying performance on these data. The proposed pipeline is not suitable to clean the RA-ILD data set and to improve the performance of the DNN in the diagnosis of ILD. The same comments expressed at the end of Section 5.1 hold also in this case.

5.4. Comparison to other data sets and works

The pre-processing pipeline presented in Section 3 has been also applied to publicly available data sets, for instance the ICBHI 2017 Challenge data set. The number of auscultations classified as bad is negligible as expected, since during the annotation process noisy signal has been probably discarded by physicians. For the sake of completeness, the performance of the techniques presented in [6,26] for the classification of lung sounds are summarized in Table 6 and compared to our results devised in Section 5.3 for the CTD-ILD data set (see Table 4). Although the comparison cannot be deemed totally fair since the data sets are different, some general indications can be devised. The pipeline proposed in [6] is composed by a feature extraction based on STFT and MFCCs and Knn classification. The achieved accuracy is 93%. The pipeline proposed in [26] exploits EMD for denoising, MFFCs and time domain parameters for feature extraction. Knn and Boosted Trees are employed for classification with an accuracy of 77% and 87.40%, respectively. We can presume that the clean data set provided by our pipeline leads our DNN to significantly outperform its counterparts.

6. Conclusions

Our findings highlight the prominent role of Knn and LogitBoost classifiers in cleaning the data set and enhancing the quality of auscultations processed by the DNN. Knn evidenced a peculiar capacity to capture local similarities and to reject outliers, so delivering a clean data set and promoting an efficient learning. The DNN designed for the diagnosis of CTD-ILD from lung sounds can provide formidable performance, namely accuracy, F1-score and F2-score of 97%. Considering that the screening of ILD in patients affected by chronic autoimmune diseases is still an open issue, the proposed DNN represents the enabling technology for raising the early diagnostic suspicion of CTD-ILD. This tool is safe, reliable and cheap. Then HRCT can be selectively prescribed, thus reducing the exposition of patients to ionizing radiation and decreasing the cost for the national health system.

CRediT authorship contribution statement

Alessandra Fava: Investigation, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Behnood Dianat:** Investigation, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Alessandro Bertacchini:** Funding acquisition, Investigation. **Andreina Manfredi:** Data curation, Validation. **Marco Sebastiani:** Data curation, Validation. **Marco Modena:** Investigation. **Fabrizio Pancaldi:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, 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.

Table 6
Comparison of the performance of the proposed processing pipeline and DNN on the CTD-ILD data set with respect to counterparts available in the literature.

Algorithm	Accuracy	Recall	F1-score	Precision	Specificity
Our DNN - Knn	97% ± 1%	96% ± 1%	97% ± 1%	98% ± 1%	98% ± 1%
Knn [6]	93% ± 1%	93% ± 1%	93% ± 1%	94% ± 1%	
Knn [26]	77% ± 2%	64% ± 2%			95% ± 2%
Our DNN - LogitBoost	91% ± 1%	95% ± 1%	91% ± 1%	89% ± 1%	85% ± 1%
Boosted trees [26]	87% ± 2%	97% ± 1%			78% ± 2%

Data availability

The data that has been used is confidential.

References

[1] F. Pancaldi, M. Sebastiani, G. Cassone, F. Luppi, S. Cerri, G.D. Casa, A. Manfredi, Analysis of pulmonary sounds for the diagnosis of interstitial lung diseases secondary to rheumatoid arthritis, Comput. Biol. Med. 96 (2018) 91–97, http://dx.doi.org/10.1016/j.compbiomed.2018.03.006.

[2] F. Pancaldi, G.S. Pezzuto, G. Cassone, M. Morelli, A. Manfredi, M. D'Arienzo, C. Vacchi, F. Savorani, G. Vinci, F. Barsotti, M.T. Mascia, C. Salvarani, M. Sebastiani, VECTOR: An algorithm for the detection of COVID-19 pneumonia from velcro-like lung sounds, Comput. Biol. Med. 142 (2022) 105220, http://dx.doi.org/10.1016/j.compbiomed.2022.105220.

[3] B. Dianat, P. La Torraca, A. Manfredi, G. Cassone, C. Vacchi, M. Sebastiani, F. Pancaldi, Classification of pulmonary sounds through deep learning for the diagnosis of interstitial lung diseases secondary to connective tissue diseases, Comput. Biol. Med. 160 (2023) 106928, http://dx.doi.org/10.1016/j.compbiomed.2023.106928.

[4] A. Manfredi, G. Cassone, S. Cerri, V. Venerito, A.L. Fedele, M. Trevisani, F. Furini, O. Addimanda, F. Pancaldi, G.D. Casa, R. D'Amico, R. Vicini, G. Sandri, P. Torricelli, I. Celentano, A. Bortoluzzi, N. Malavolta, R. Meliconi, F. Iannone, E. Gremese, F. Luppi, C. Salvarani, M. Sebastiani, Diagnostic accuracy of a velcro sound detector (VECTOR) for interstitial lung disease in rheumatoid arthritis patients: the InSPIRAtE validation study (INterStitial pneumonia in rheumatoid ArThritis with an electronic device), BMC Pulm. Med. 19 (1) (2019) http://dx.doi.org/10.1186/s12890-019-0875-x.

[5] A. Manfredi, M. Sebastiani, S. Cerri, C. Vacchi, R. Tonelli, G.D. Casa, G. Cassone, A. Spinella, P. Fabrizio, F. Luppi, C. Salvarani, Acute exacerbation of interstitial lung diseases secondary to systemic rheumatic diseases: a prospective study and review of the literature, J. Thorac. Dis. 11 (4) (2019) 1621–1628, http://dx.doi.org/10.21037/jtd.2019.03.28.

[6] A. Ullah, M.S. Khan, M.U. Khan, F. Mujahid, Automatic classification of lung sounds using machine learning algorithms, in: 2021 International Conference on Frontiers of Information Technology, FIT, 2021, pp. 131–136, http://dx.doi.org/10.1109/FIT53504.2021.00033.

[7] S. Gupta, M. Agrawal, D. Deepak, Gammatonegram based triple classification of lung sounds using deep convolutional neural network with transfer learning, Biomed. Signal Process. Control 70 (2021) 102947, http://dx.doi.org/10.1016/j.bspc.2021.102947.

[8] N. Sengupta, M. Sahidullah, G. Saha, Lung sound classification using cepstral-based statistical features, Comput. Biol. Med. 75 (2016) 118–129, http://dx.doi.org/10.1016/j.compbiomed.2016.05.013.

[9] G. Altan, Y. Kutlu, A. Gökçen, Chronic obstructive pulmonary disease severity analysis using deep learning on multi-channel lung sounds, Turk. J. Electr. Eng. Comput. Sci. 28 (2020) 2979–2996, http://dx.doi.org/10.3906/elk-2004-68.

[10] G. Altan, Y. Kutlu, A.Ö. Pekmezci, S. Nural, Deep learning with 3D-second order difference plot on respiratory sounds, Biomed. Signal Process. Control 45 (2018) 58–69, http://dx.doi.org/10.1016/j.bspc.2018.05.014, URL https://www.sciencedirect.com/science/article/pii/S1746809418301162.

[11] International conference on biomedical health informatics challenge data set, 2017, URL https://bhichallenge.med.auth.gr/ICBHI_2017_Challenge.

[12] B. Rocha, et al., A respiratory sound database for the development of automated classification, in: International Conference on Biomedical and Health Informatics, Springer, 2018, pp. 33–37, http://dx.doi.org/10.1007/978-981-10-7419-6-6.

[13] A. Manzoor, Q. Pan, H.J. Khan, S. Siddeeq, H.M.A. Bhatti, M.A. Wedagu, Analysis and detection of lung sounds anomalies based on NMA-RNN, in: 2020 IEEE International Conference on Bioinformatics and Biomedicine, BIBM, 2020, pp. 2498–2504, http://dx.doi.org/10.1109/BIBM49941.2020.9313197.

[14] L. Fraiwan, O. Hassanin, M. Fraiwan, B. Khassawneh, A.M. Ibnian, M. Alkhodari, Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers, Biocybern. Biomed. Eng. 41 (1) (2021) 1–14, http://dx.doi.org/10.1016/j.bbe.2020.11.003, URL https://www.sciencedirect.com/science/article/pii/S0208521620301297.

[15] J. Gnitecki, Z.M. Moussavi, Separating heart sounds from lung sounds, IEEE Eng. Med. Biol. Mag. 26 (1) (2007) 20.

[16] J. Gnitecki, Z. Moussavi, H. Pasterkamp, Recursive least squares adaptive noise cancellation filtering for heart sound reduction in lung sounds recordings, in: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439), Vol. 3, 2003, pp. 2416–2419, http://dx.doi.org/10.1109/IEMBS.2003.1280403, Vol.3.

[17] V.K. Iyer, P.A. Ramamoorthy, H. Fan, Y. Ploysongsang, Reduction of heart sounds from lung sounds by adaptive filterng, IEEE Trans. Biomed. Eng. BME-33 (12) (1986) 1141–1148, http://dx.doi.org/10.1109/TBME.1986.325693.

[18] S.A. Baharanchi, M. Vali, M. Modaresi, Noise reduction of lung sounds based on singular spectrum analysis combined with discrete cosine transform, Appl. Acoust. 199 (2022) 109005.

[19] I. Hossain, Z. Moussavi, An overview of heart-noise reduction of lung sound using wavelet transform based filter, in: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439), Vol. 1, 2003, pp. 458–461, http://dx.doi.org/10.1109/IEMBS.2003.1279719, Vol.1.

[20] L.J. Hadjileontiadis, S.M. Panas, A wavelet-based reduction of heart sound noise from lung sounds, Int. J. Med. Inform. 52 (1–3) (1998) 183–190.

[21] G.-C. Chang, Y.-P. Cheng, Investigation of noise effect on lung sound recognition, in: 2008 International Conference on Machine Learning and Cybernetics, Vol. 3, 2008, pp. 1298–1301, http://dx.doi.org/10.1109/ICMLC.2008.4620605.

[22] Y.P. Kahya, E.Ç. Güler, B. Sankur, T. Engin, Detection and clustering analysis of crackles in respiratory sounds, in: 1992 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol. 6, 1992, pp. 2527–2528, http://dx.doi.org/10.1109/IEMBS.1992.5761571.

[23] D. Emmanouilidou, E.D. McCollum, D.E. Park, M. Elhilali, Adaptive noise suppression of pediatric lung auscultations with real applications to noisy clinical settings in developing countries, IEEE Trans. Biomed. Eng. 62 (9) (2015) 2279–2288.

[24] B. Sangeetha, R. Periyasamy, Performance metrics analysis of adaptive threshold empirical mode decomposition denoising method for suppression of noise in lung sounds, in: 2021 Seventh International Conference on Bio Signals, Images, and Instrumentation, ICBSII, IEEE, 2021, pp. 1–6.

[25] M. Pourazad, Z. Moussavi, F. Farahmand, R. Ward, Heart sounds separation from lung sounds using independent component analysis, in: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, IEEE, 2006, pp. 2736–2739.

[26] S. Aziz, M.U. Khan, M. Shakeel, Z. Mushtaq, A.Z. Khan, An automated system towards diagnosis of pneumonia using pulmonary auscultations, in: 2019 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics, MACS, 2019, pp. 1–7, http://dx.doi.org/10.1109/MACS48846.2019.9024789.

[27] K. Dragomiretskiy, D. Zosso, Variational mode decomposition, IEEE Trans. Signal Process. 62 (3) (2014) 531–544, http://dx.doi.org/10.1109/tsp.2013.2288675.

[28] J. Driedger, M. Müller, S. Disch, Extending harmonic-percussive separation of audio signals, in: Proceedings of the 15th International Conference on Music Information Retrieval (ISMIR), Taipei, Taiwan, 2014, pp. 611–616.

[29] K. Taunk, S. De, S. Verma, A. Swetapadma, A brief review of nearest neighbor algorithm for learning and classification, in: 2019 International Conference on Intelligent Computing and Control Systems, ICCS, 2019, pp. 1255–1260, http://dx.doi.org/10.1109/ICCS45141.2019.9065747.

[30] S. Safavian, D. Landgrebe, A survey of decision tree classifier methodology, IEEE Trans. Syst. Man Cybern. 21 (3) (1991) 660–674, http://dx.doi.org/10.1109/21.97458.

[31] J. Friedman, T. Hastie, R. Tibshirani, Additive logistic regression: A statistical view of boosting, Ann. Statist. 28 (2000) 337–407, http://dx.doi.org/10.1214/aos/1016218223.

[32] K.P. Murphy, et al., Naive bayes classifiers, Univ. B. C. 18 (60) (2006) 1–8.

[33] G. Altan, Y. Kutlu, RespiratoryDatabase@TR (COPD severity analysis), 2020, http://dx.doi.org/10.17632/p9z4h98s6j.1, URL https://data.mendeley.com/datasets/p9z4h98s6j/1.

[34] D. Emmanouilidou, M. Elhilal, Characterization of noise contaminations in lung sound recordings, in: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, 2013, pp. 2551–2554, http://dx.doi.org/10.1109/EMBC.2013.6610060.

[35] A. Leal, R. Couceiro, I. Chouvarda, N. Maglaveras, J. Henriques, R. Paiva, P. Carvalho, C. Teixeira, Detection of different types of noise in lung sounds, in: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, 2016, pp. 5977–5980, http://dx.doi.org/10.1109/EMBC.2016.7592090.

[36] A. Manfredi, G. Cassone, C. Vacchi, F. Pancaldi, G.D. Casa, S. Cerri, L.D. Pasquale, F. Luppi, C. Salvarani, M. Sebastiani, Usefulness of digital velcro crackles detection in identification of interstitial lung disease in patients with connective tissue diseases, Arch. Rheumatol. (2020) http://dx.doi.org/10.46497/archrheumatol.2021.7975.

[37] M. Sebastiani, C. Vacchi, G. Cassone, F. Atzeni, M. Biggioggero, A. Carriero, G.L. Erre, A.L. Fedele, F. Furini, P. Tomietto, V. Venerito, B. Atienza-Mateo, G.D. Casa, S. Cerri, G. Sandri, A. Palermo, E. Galli, F. Pancaldi, M.A. González-Gay, C. Salvarani, A. Manfredi, THU0150 interstitial lung disease related to rheumatoid arthritis. What do we don't know? the LIRA study (lung involvement in rheumatoid arthritis)., Ann. Rheum. Dis. 79 (Suppl 1) (2020) 290–291, http://dx.doi.org/10.1136/annrheumdis-2020-eular.3516.

[38] H. Akima, A new method of interpolation and smooth curve fitting based on local procedures, J. ACM 17 (4) (1970) 589–602, http://dx.doi.org/10.1145/321607.321609.

[39] B. McFee, C. Raffel, D. Liang, D. Ellis, M. McVicar, E. Battenberg, O. Nieto, Librosa: Audio and music signal analysis in Python, in: Proceedings of the 14th Python in Science Conference, SciPy, 2015, http://dx.doi.org/10.25080/majora-7b98e3ed-003.

[40] C.R. Harris, K.J. Millman, S.J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N.J. Smith, R. Kern, M. Picus, S. Hoyer, M.H. van Kerkwijk, M. Brett, A. Haldane, J.F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T.E. Oliphant, Array programming with NumPy, Nature 585 (7825) (2020) 357–362, http://dx.doi.org/10.1038/s41586-020-2649-2.

[41] W. McKinney, et al., Data structures for statistical computing in python, in: Proceedings of the 9th Python in Science Conference, Vol. 445, Austin, TX, 2010, pp. 51–56, http://dx.doi.org/10.25080/Majora-92bf1922-00a.

[42] V. Bajaj, R.B. Pachori, Classification of seizure and nonseizure EEG signals using empirical mode decomposition, IEEE Trans. Inf. Technol. Biomed. 16 (6) (2012) 1135–1142, http://dx.doi.org/10.1109/titb.2011.2181403.

[43] C.M. Sweeney-Reed, S.J. Nasuto, A novel approach to the detection of synchronisation in EEG based on empirical mode decomposition, J. Comput. Neurosci. 23 (1) (2007) 79–111, http://dx.doi.org/10.1007/s10827-007-0020-3.

[44] R.J. Martis, U.R. Acharya, J.H. Tan, A. Petznick, R. Yanti, C.K. Chua, E.Y.K. Ng, L. Tong, Application of empirical mode decomposition (EMD) for automated detection of epilepsy using EEG signals, Int. J. Neural Syst. 22 (06) (2012) 1250027, http://dx.doi.org/10.1142/s012906571250027x.

[45] D. Zosso, Variational mode decomposition, 2013, URL https://www.mathworks.com/matlabcentral/fileexchange/44765-variational-mode-decomposition.

[46] M. Nazari, S.M. Sakhaei, Variational mode extraction: A new efficient method to derive respiratory signals from ECG, IEEE J. Biomed. Health Inform. 22 (4) (2018) 1059–1067, http://dx.doi.org/10.1109/jbhi.2017.2734074.

[47] Q. Zhou, J. Shan, W. Ding, C. Wang, S. Yuan, F. Sun, H. Li, B. Fang, Cough recognition based on mel-spectrogram and convolutional neural network, Front. Robot. AI 8 (2021) http://dx.doi.org/10.3389/frobt.2021.580080.

[48] L. Rabiner, R. Schafer, Theory and Applications of Digital Speech Processing, Prentice Hall Press, 2010.

[49] 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, http://dx.doi.org/10.1109/JBHI.2023.3248265.

[50] S.N. Levine, J.O. Smith III, A sines+ transients+ noise audio representation for data compression and time/pitch scale modifications, in: Audio Engineering Society Convention 105, Audio Engineering Society, 1998.

[51] A. Petrovsky, E. Azarov, A. Petrovsky, Hybrid signal decomposition based on instantaneous harmonic parameters and perceptually motivated wavelet packets for scalable audio coding, Signal Process. 91 (6) (2011) 1489–1504, http://dx.doi.org/10.1016/j.sigpro.2010.09.005.

[52] R. Riella, P. Nohama, J. Maia, Method for automatic detection of wheezing in lung sounds, Braz. J. Med. Biol. Res. 42 (7) (2009) 674–684, http://dx.doi.org/10.1590/S0100-879X2009000700013.

[53] A. Rizal, W. Priharti, D. Rahmawati, H. Mukhtar, Classification of pulmonary crackle and normal lung sound using spectrogram and support vector machine, J. Biomim. Biomater. Biomed. Eng. 55 (2022) 143–153, http://dx.doi.org/10.4028/p-tf63b7.

[54] A. Parkhi, M. Pawar, Analysis of deformities in lung using short time Fourier transform spectrogram analysis on lung sound, in: 2011 International Conference on Computational Intelligence and Communication Networks, 2011, pp. 177–181, http://dx.doi.org/10.1109/CICN.2011.35.

[55] W. Zhang, F. Zeng, X. Wu, X. Zhang, R. Jiang, A comparative study of ensemble learning approaches in the classification of breast cancer metastasis, in: 2009 International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing, 2009, pp. 242–245, http://dx.doi.org/10.1109/IJCBS.2009.23.

[56] G. Altan, Y. Kutlu, Y. Garbİ, A.Ö. Pekmezcİ, S. Nural, Multimedia respiratory database (RespiratoryDatabase@TR): Auscultation sounds and chest X-rays, Natl. Eng. Sci. 2 (3) (2017) 59–72, http://dx.doi.org/10.28978/nesciences.349282.