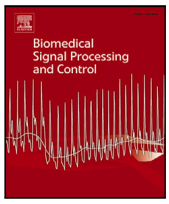
Biomedical Signal Processing and Control 119 (2026) 109864

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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



Classification of lung sounds for the detection of interstitial lung disease secondary to rheumatoid arthritis

Fabrizio Pancaldi *, Luca Dibiase

Department of Sciences and Method for Engineering (DISMI), University of Modena and Reggio Emilia, Reggio Emilia, Italy

ARTICLE INFO

Keywords:
Digital signal processing
Deep learning
Lung sounds
Interstitial lung disease
Rheumatoid arthritis

ABSTRACT

Rheumatoid arthritis is an autoimmune disease impacting around 1% of population. One of the most severe comorbidity is interstitial lung disease. Currently, the treatment is effective only in the very early stages of the disease. Symptoms appear late in the clinical history and are not useful for diagnosis. A routine use of high resolution computer tomography for screening programs is not advisable for both exposition to ionizing radiation of patients and high costs to be sustained by the national health system. Lung auscultation can reveal the so called "velcro crackle" associated to different radiological patterns. In this work we propose a new pipeline for pre-processing and classification of lung sounds suitable to the detection of interstitial lung disease in patient affected by rheumatoid arthritis. The data set has been collected in a clinical study at the university hospital of Modena (Italy). Ground truth is represented by the high resolution computer tomography report. Accuracy and F1-score of our solution are 83.2% and 77,9% in the classification of lung sounds, respectively. Combining the predictions of the classifier for distinct auscultations of the same patient, the accuracy and F1-score get as high as 87.8% and 87,1%, respectively. Considering that physical lung auscultation is safe for the patient and cheap for the national health system, the proposed solution can pave the way for a screening campaign aimed at the early detection of interstitial lung disease secondary to rheumatoid arthritis.

1. Introduction

Rheumatoid arthritis (RA) is an autoimmune disease impacting around 1% of population [1]. The first symptom to appear is swelling of the joints; then the disease progressively leads to pain, chronic pain and even deformity of joints. The most severe comorbidities are cardiovascular disease and interstitial lung disease (ILD). Posterior statistics and retrospective studies have shown that the life expectation of patients affected by ILD secondary to RA (RA-ILD) is very poor, on the order of 3–8 years [2,3]. In practice, the lung parenchyma is replaced by fibrotic tissue with a reduced capability to exchange oxygen and carbon dioxide between blood and air. Currently, one drug is capable to stop the progression of the ILD, but only if the treatment is dispensed in the very early stages of the disease. The pathogenesis of RA and RA-ILD is almost unknown, as well as the onset of RA-ILD is not predictable. Symptoms appear late in the clinical history and are not useful for diagnosis. In fact, patients can be asymptomatic at the early stages of the disease, and 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 exposure to ionizing radiation of patients. To improve the prescriptive appropriateness of HRCT for the early diagnosis of ILD, 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. Velcro crackles generated in the lung parenchyma has been independently associated with different radiological patterns, namely honeycombing, ground glass and traction bronchiectasis [4].

The detection and/or classification of abnormal lung sounds represents a well consolidated field of research. On the one hand, several works aim at the classification of lung sounds collected in heterogeneous data sets. We mean heterogeneous data set a collection of lung sounds coming from distinct sources that have not been systematically recorded with the same formality. The ICBHI 2017 data set [5] consists of 6898 respiratory cycles, of which 1864 are labeled as crackles, 886 are classified as wheezes, and 506 belong to both crackles and wheezes classes, for a total of 920 audio samples recorded from 126 subjects. The data set of [6] is composed by 532 samples of which

* Corresponding author.
E-mail address: fpancaldi@unimore.it (F. Pancaldi).

https://doi.org/10.1016/j.bspc.2026.109864
Received 5 July 2025; Received in revised form 29 December 2025; Accepted 13 February 2026
Available online 20 February 2026

60 healthy and 472 abnormal with crepitation or wheezing. The so called HF_Lung_V1 includes 9765 audio files [7], with 34 095 labels from inhalation, 18 349 labels from exhalation, 13 883 labels are listed as continuous adventitious sound (8457 wheeze sounds, 686 samples of stridor, 4740 rhonchi), and 15 606 are discontinuous adventitious sounds, all of which are crackles. Shi et al. proposed a pipeline based on a combination of wavelet coefficients, high-pass filtering and Mel spectrogram with transfer learning on VGGish model and a bidirectional gated recurrent unit neural network [8]. The data set was collected in hospital and is composed by 120 normal lung sounds, 156 pneumonia sounds and 108 asthma sounds suitably selected among those of the better quality. The accuracy was of 87.4%. Phetton et al. [9] worked on abnormal lung sounds characterized by crackles or wheezing. Spectrograms provided by the short-time Fourier transform (STFT) were analyzed through the convolutional neural network (CNN) GoogleNet achieving an accuracy of 85.3%. Majzoobi et al. [10] focused of classifying asthma and chronic obstructive pulmonary disease (COPD) with an hybrid neural network. Long-short term memory (LSTM) blocks were embedded in a CNN to achieve an accuracy of 97.4%. Roy and Satija investigated the identification of ILD from respiratory sounds through the use of a dedicated framework [11]. Pre-processing is based on band-pass Butterworth filtering, sample cutting and z-score normalization. Classification relies on a new sinc convolution-based residual CNN characterized by 48 914 parameters. The new architecture named ILDNet has been tested on a combination of the two public data sets BRACET [12] and KAUH [13]. The accuracy, sensitivity and specificity are of 81,3%, 78,9% and 83,3%, respectively. Roy and Satija devised also new frameworks for the detection of COPD [14,15]. The work [14] is based on the YAMNet, a model pre-trained on the AudioSet data set [16]. The proposed solution achieved an accuracy of 99.3% on the RespiratoryDatabase@TR [17]. The approach [15] exploits multiple time-frequency representations of respiratory sounds, namely Mel spectrogram, constant-Q transform and Mel frequency cepstral coefficients. The devised multi-head self-organized operational neural network has been tested on three publicly available data set including the ICBHI 2017 [5], the Chest Wall Lung Sound Database [13] and the RespiratoryDatabase@TR [17]. The proposed solution can provide an accuracy as high as 99,8%. Roy and Satija worked also on the classification of adventitious respiratory sounds correlated to various lung disorders [18]. The approach is based on Mel spectrogram chunks feeding different pre-trained audio neural networks like VGGish, YamNet and OpenL3. Performance assessment on the public ICBHI 2017 data set [5] lead to sensitivity and specificity of 79.6% and 82.7%, respectively.

It is worth pointing out that in all these works the ground truth is represented by subjective annotations of physicians or general doctors. On the other hand, few works deals with the detection of pathological lung sounds included in a clinical study where the ground truth is given by the HRCT report of the radiologist. Pancaldi et al. [19] proposed a processing pipeline based of time-frequency analysis through STFT, identification of the inspiration period, bandwidth computation and hard thresholding. The overall accuracy over a RA-ILD data set of 137 patients was 83.9% [20]. Manfredi et al. [21] employed an algorithmic approach to the diagnosis of ILD secondary to connective tissue diseases (CTD). The idea behind the work [21] consists of considering velcro crackles as voiced/unvoiced sounds and filtering them through linear predictive coding. Principal component analysis and hard thresholding is used for classification. The flow chart of the processing chain is described in detail in [22]. The overall diagnostic accuracy was 82.6%. Dianat et al. [23] designed a pipeline for pre-processing lung sounds collected from patients affected by CTD-ILD. Variational mode decomposition (VMD) [24] is exploited to highlight pathological lung sounds still suppressing background noise and artifacts evidenced in "bad" auscultations. Mel filterbank is used for time-frequency analysis. Various deep neural network (DNN) architectures are adopted for the classification of Mel spectrograms. The overall diagnostic accuracy for CTD-ILD was 91%. Fava et al. [25] investigated an algorithmic approach to clean the data sets of lung sounds from "bad" auscultations. We mean for bad auscultation a lung sounds where the useful respiratory signal is absent or is overwhelmed by noise or artifacts. Pre-processing is based on VMD and harmonic-percussive source separation (HPSS) [26]. Classification into good and bad signal is performed through the k-nearest neighbors algorithm. The overall accuracy of the DNN developed in [23] applied to the clean data set in the diagnosis of CTD-ILD and RA-ILD is 97% and 68%, respectively. Unlike the impressive performance of the DNN proposed in [23,25] for the diagnosis of CTD-ILD, the overall accuracy of the same DNN for the diagnosis of RA-ILD can be deemed quite poor.

The scope of this work consists of designing a new pipeline capable to raise the diagnostic accuracy for RA-ILD at a level suitable to be employed in the clinical practice. The proposed pipeline is based on high pass filtering and variational mode extraction (VME) [27] for signal denoising. Sample cutting is adopted as a simple and easy approach to data augmentation. STFT and HPSS are employed for time-frequency analysis and to highlight harmonic components related to pathological lung sounds, respectively. The HPSS filtered spectrograms feed a CNN for binary classification. The CNN is trained through transfer learning. The performance of the pipeline is assessed with respect to its capability in the classification of single auscultations and patients. The two output classes include elements, i.e. auscultations or patients, either positive or negative to RA-ILD. The ground truth is represented by the HRCT report.

The remaining of this paper is organized as follows. The clinical study is described in Section 2. The developed pipeline is explained in Section 3. Numerical results are presented in Section 4. Some conclusions are offered in Section 5.

2. Clinical study

The data set considered in this study is described in detail in [20]. In this Section we recall some information relevant to the scope of this work. The data set includes 137 RA patients classified according to 1987 or 2010 American College of Rheumatology criteria [28,29]. All consecutive RA patients with a recent HRCT evaluation were eligible for the study and enrolled in a six-month period. According to clinical history, HRCT should have been performed within 12 months in the absence of the subsequent appearance or variation of signs or symptoms suggestive of lung disease (cough, dyspnoea, velcro sound at routine clinical examination). Exclusion criteria were: (a) significant variations in respiratory symptoms after HRCT imaging (when possible, a new HRCT was requested); (b) presence of pleural effusion or pneumothorax at HRCT; (c) a diagnosis overlapping with CTD. All HRCT images were transferred on DICOM format, anonymized, coded and evaluated in a blind manner by an expert thoracic radiologist for the assessment of ILD. Respiratory sounds were recorded in 3 pulmonary fields bilaterally (1 at the basal field in paravertebral position, 1 at the basal field in axillary position, 1 at the middle field in paravertebral position), for a total of 6 auscultations per patient. Lung were acquired in a silent environment through the commercial electronic stethoscope Littmann 3200. The sampling frequency is $f_s = 4$ kHz. Audio files are saved in .wav format. The study [20] was approved by the ethical committee of Modena (number 2636, July 9, 2014, Italy) and all patients signed an informed consent form.

The radiologists' reports represent the ground truth for this work; 78 patients were classified as negative to ILD and 59 as positive to ILD. Other relevant clinical features extracted from the study [20] are shown in Table 1, namely number of smokers, male/female proportion, mean age of patients, rheumatoid factor and forced vital capacity.

3. Pipeline for data pre-processing and deep learning

The pipeline developed in this study is shown in Fig. 1. The acquired signal is firstly high-pass filtered to suppress low frequency components that do not carry useful information about pathological

Table 1
Clinical features of our data set extracted from the study [20].

	Total	Negative	Positive
Number	137	78	59
Smoker	51	31	20
Sex M/F	48/89	22/56	26/33
Mean age (years)	$56,1 \pm 12,9$	$55,3 \pm 12,3$	$57,0 \pm 13,6$
Rheumatoid factor (%)	78,6%	81,1%	75,4%
Forced vital capacity	$91,8\% \pm 22,3$	$93,1\% \pm 21,3$	$90,3\% \pm 23,7$

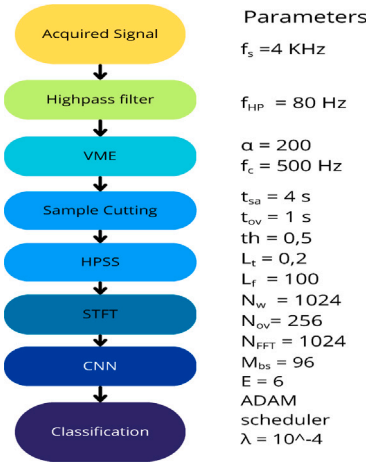


Fig. 1. Pipeline for the classification of lung sounds.

sounds, like for instance heartbeat and artifacts related to stethoscope handling (rubbing the diaphragm on the skin, tapping fingers on the case, pushing buttons, changing grip, ...). This approach was suggested in [23,25]. We denote with $f_{HP} = 80$ Hz the cutting frequency of the high-pass filter of the first order. VME [27] has been employed for signal decomposition, i.e. for separating components that evidence a given mathematical structure. VME has been introduced in [27] to overcome some issues of VMD [24]. In fact, VME attempts the separation of a single compact signal around a trial center frequency f_c, rather than concurrently decomposing a given number of modes as in VMD. The scope of this step consists of further denoising the signal and focusing as much as possible on respiratory sounds useful for diagnosis. Under the VME approach, the original signal $s(t)$ is represented as the superposition of the desired mode $u_d(t)$ and the residual signal $s_r(t)$, i.e.

$$s(t) = u_d(t) + s_r(t). \tag{1}$$

The desired mode takes the form of an amplitude-modulated-frequency-modulated signal as

$$u_d(t) = A_d(t)\cos\left(\varphi_d(t)\right) \tag{2}$$

where $A_d(t)$ is the envelope and $\varphi_d(t)$ is the phase. The desired mode $u_d(t)$ shall evidence two properties. Firstly it shall be compact around its center frequency f_c. Consequently, it shall minimize the metric

$$J_1 = \left\| \partial t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_d(t) \right] * \exp\left(-j\omega_d t\right) \right\|_2^2, \tag{3}$$

where $\delta(t)$ is the Dirac distribution, $*$ denotes convolution, $\|\cdot\|_2$ denotes the L^2 norm and $\omega_d = 2\pi f_c$. Secondly, the spectral overlap of the residual signal $s_r(t)$ with the desired mode $u_d(t)$ should be minimized, i.e. the energy of the residual signal should be minimized at frequency band where the desired mode lies. Practically, a filter with impulse response $\beta(t)$ and frequency response $\bar{\beta}(\omega)$ is considered to guarantee that the energy of $s_r(t)$ at ω_d is zero. This constraint leads to the penalty function

$$J_2 = \left\| \beta(t) * s_r(t) \right\|_2^2. \tag{4}$$

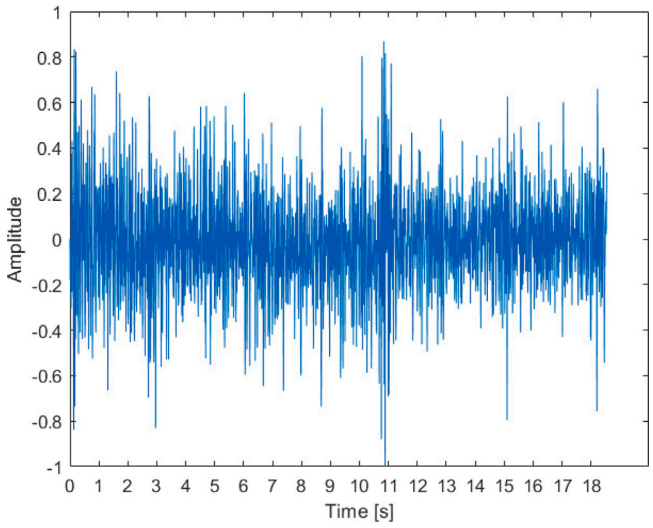


Fig. 2. Signal acquired from an RA patient negative to ILD.

The work [27] suggests the use of the filter characterized by frequency response

$$\bar{\beta}(\omega) = \frac{1}{\alpha\left(\omega - \omega_d\right)^2}. \tag{5}$$

Hence, the problem of finding the desired mode can be expressed as the minimization problem

$$\min_{u_d, \omega_d, s_r} \left\{ \alpha J_1 + J_2 \right\} \tag{6}$$

subject to Eqs. (1) and (2). The parameter α practically controls the compactness of the extracted mode. Figs. 2 and 3 show two examples of acquired signals negative and positive to ILD, respectively, whereas Figs. 4 and 5 show the related components after high-pass filtering and VME. The effectiveness of this approach is witnessed by the signals reported in Fig. 5 sounding at the human ears as the so called velcro crackles.

Sample cutting is performed on the basis of frames lasting a time of $t_{sa} = 4$ s with an overlapping of $t_{ov} = 1$ s, as suggested in [23]. This approach allows to extend the data set for classification and consequently to minimize overfitting. HPSS [26] is exploited to suppress the percussive component and to highlight the harmonic component, since some residual noise may remain from strong impulses, like for instance coughing fits. We denote with $th = 0.5$, $L_t = 0.2$ s and $L_f = 100$ Hz the threshold, the length of the median filter along time and the length of the median filter along frequency, respectively. Time-frequency analysis is performed through STFT mainly because of its close affinity to HPSS. The window length is $N_w = 1024$ samples, the number of overlapping samples is $N_{ov} = 256$ and the order of the fast Fourier transform (FFT) is $N_{FFT} = 1024$. Spectrograms are resized to images of 224 × 224 pixels in RGB format with depth of 8 bits.

The CNN employed in this work is Googlenet [30]. It is characterized by 22 layers, a mini batch size $M_{bs} = 96$ and more than 7 million of parameters. ReLu rectifier is used as activation function. The CNN is pre-trained with the widely adopted ImageNet data set [31] including more than 14 millions of images labeled in about 20.000 classes. Then, transfer learning is applied exploiting the experimental data set described in Section 2. To this aim, the last layer of the Googlenet including the ImageNet classes is substituted with a blank layer characterized by 2 classes (negative and positive to RA-ILD) before training. Classification is based of images, i.e. spectrograms, rather than lung sounds for two main reasons. Firstly, the data sets of images are by far larger than the audio counterparts, so transfer learning from a limited data set like ours can be more effective.

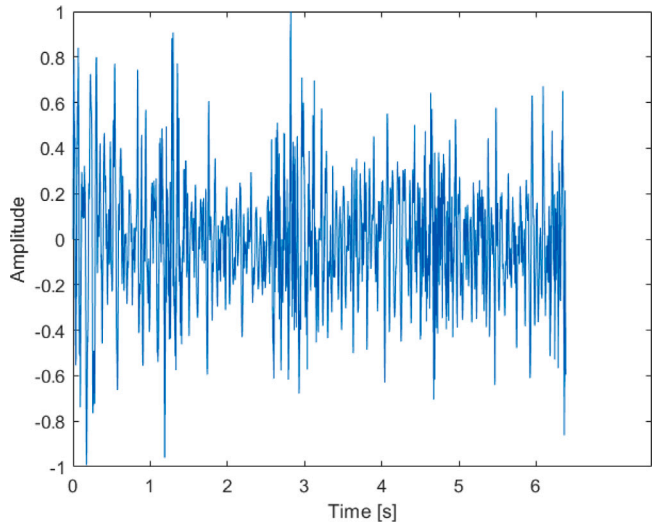


Fig. 3. Signal acquired from an RA patient positive to ILD.

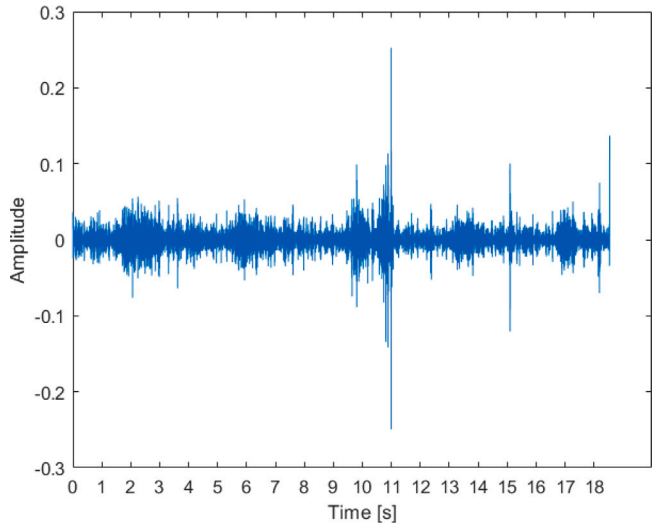


Fig. 4. Signal related to an RA patient negative to ILD after high-pass filtering and VME.

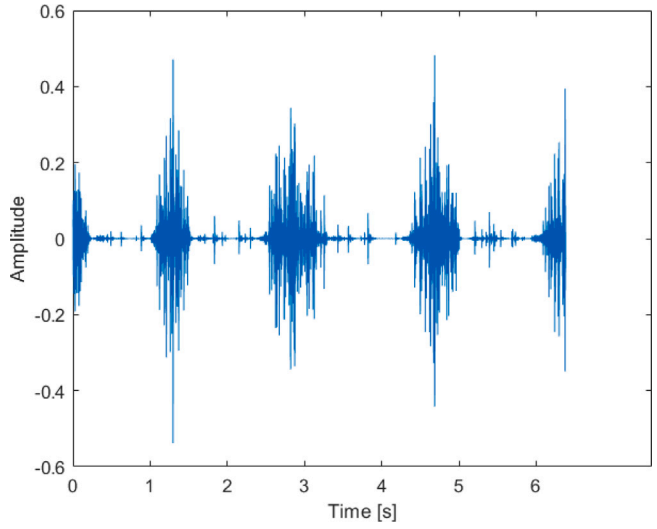


Fig. 5. Signal related to an RA patient positive to ILD after high-pass filtering and VME.

Secondly, we deem that time-frequency analysis can highlight and steer the deep neural network to the features that are more relevant for the diagnosis of velcro crackles. These thoughts are corroborated by the results of the work [18], where various pre-trained audio neural networks cannot achieve sensitivity and specificity larger than of 80% and 83%, respectively, when other pipelines can achieve an accuracy of about 99% on the ICBHI 2017 data set. Furthermore VGGish and YAMNet are tied to Mel Spectrogram, so limiting the options on time-frequency analysis. ADAM scheduler is adopted with an initial learning rate $\lambda = 10^{-4}$. The data set described in Section 2 is divided into training, validation and test set composed by the 75%, 20% and 5% of elements, respectively. The modest data imbalance is handled through random splitting with stratification sampling; this technique allows to preserve the proportion of negative and positive patients over a pool of 2024 signals. The number of epochs for training is 6; for each epoch, 15 iterations are performed to train the net. The number of epochs for training is small as a consequence of transfer learning and it has been empirically adjusted by monitoring the loss curve and selecting the most suitable parameter to prevent overfitting. The dropout is set to 40% and the corresponding validation accuracy is 67.4%, as shown in Fig. 6. The limited validation accuracy can be explained by the complexity in the classification of spectrograms. Nonetheless, overfitting is prevented.

The STFT of two auscultations acquired from a patient negative and positive to RA-ILD are shown in Figs. 7 and 8, respectively. In both Figures, the periodic alternation between inspiration and expiration breath cycles are observable. On the contrary, the spectra are quite different. In fact, considering the expiration periods, the band of the negative auscultation is approximately 80–300 Hz, whereas the band of the positive auscultation is approximately 300–800 Hz. From a qualitative point of view, the harmonic components in the band 300–800 Hz of the positive auscultation correspond to the so called velcro crackles, which is by now widely considered as an early marker of ILD.

4. Numerical results

The complexity and toughness of the data set described in Section 2 deserve a first discussion. Taking inspiration from [14], the two-dimensional t-distributed stochastic neighbor embedding (t-SNE) visualization [32] of the original lung sounds and of the harmonic components at the output of the HPSS are shown in Figs. 9 and 10, respectively. Fig. 9 evidences the randomness of the data set, since no cluster can be easily identified, as well as patients positive and negative to RA-ILD are widely mixed. The proposed pre-processing techniques play an important role in facilitating classification, in fact small clusters can be identified in Fig. 10. Nonetheless, patients positive and negative to RA-ILD are still mixed. This behavior can be easily interpreted. In many data sets like for instance the RespiratoryDatabase@TR [17], healthy auscultation records are selected among volunteers who have never used cigarettes or tobacco products, as well as have no diagnosed chronic lung history. Despite this, in our data set all the patients are affected by RA and present comorbidities, so none of them can be defined “healthy” in a strict sense.

The data set is described in Section 2 and further commented in Section 3. The parameters relevant for the proposed pipeline are introduced in Section 3 and summarized in Fig. 1. The confusion matrix of the proposed pipeline for the classification of the test set is shown in Fig. 11, whereas the related metrics are summarized in Table 2. TP, TN, FP and FN denote true positives, true negatives, false positives and false negatives, respectively. Firstly, it is worth pointing out that, generally speaking, patients negative to ILD can take more deep breaths than patients positive to ILD. Consequently, the data set in terms of samples of lung sounds is more unbalanced towards negative elements than the data set expressed in terms of patients (see Section 2). In the considered clinical problem, the number of FN should be minimized,

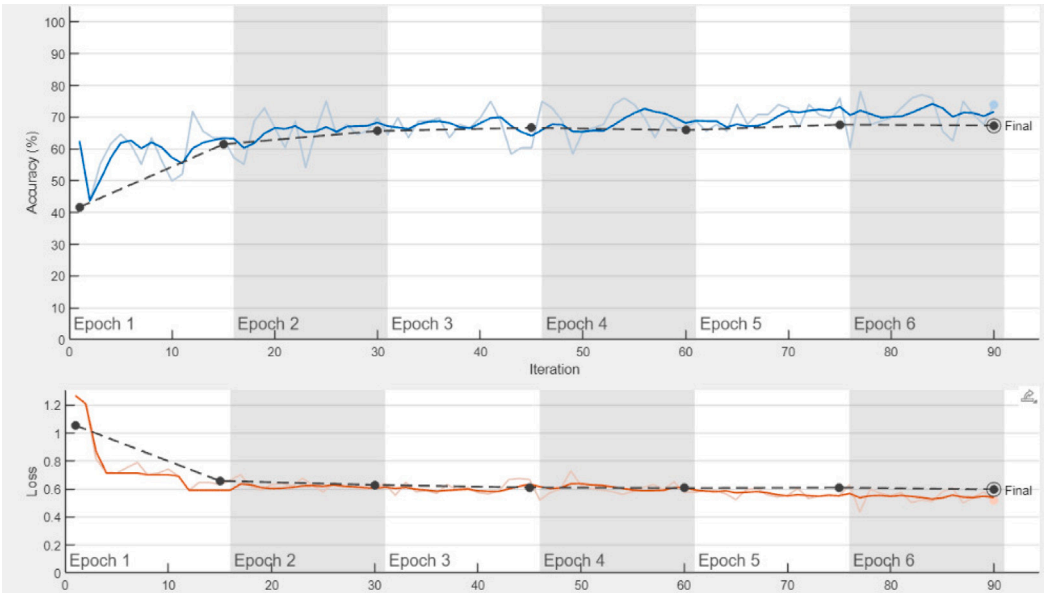


Fig. 6. Training curves in terms of accuracy and loss. Blue solid line denotes training accuracy, orange solid line is the training loss, black dotted lines represent the corresponding validation accuracy and loss.

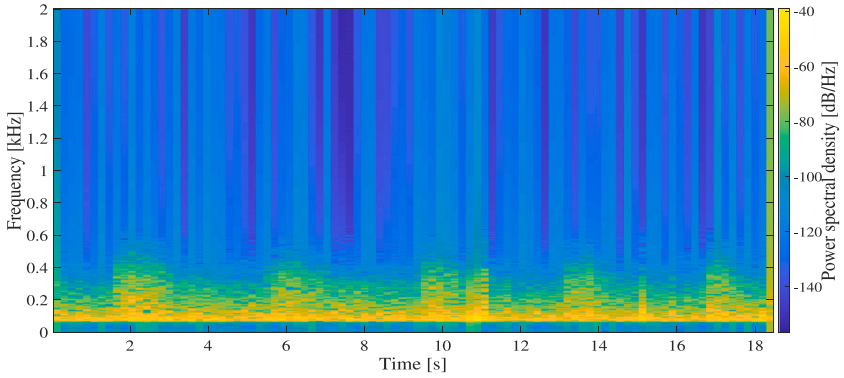


Fig. 7. STFT of an auscultation acquired from a patient negative to RA-ILD.

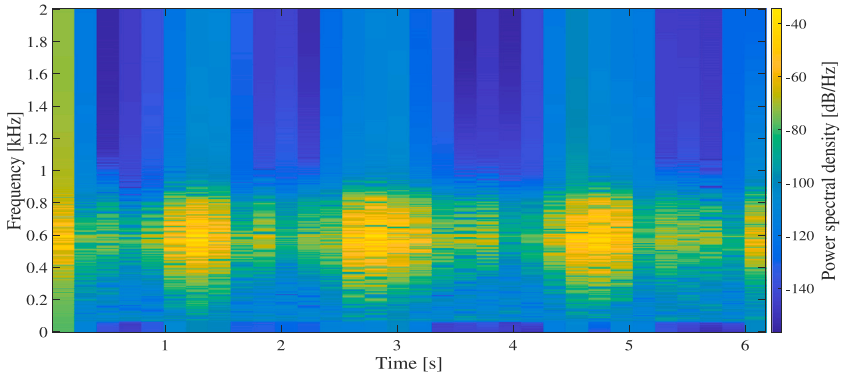


Fig. 8. STFT of an auscultation acquired from a patient positive to RA-ILD.

since this number refers to patient affected by RA that are not undergoing the proper followup for ILD. However, patients in advanced stage of ILD cannot breath deeply and the fibrotic pulmonary tissue is not "stimulated", so pathological lung sounds are not generated and cannot be detected. This turn in a fair sensibility of 73.2%. The number of FP should be minimized as well, since this number refers to patients undergoing an unnecessary HRCT. The proposed pipeline has evidenced a very good specificity of 90%. Negative and positive predictive values are quite balanced and are equal to 83.1% and 83.3%, respectively. The resulting overall accuracy is 83.2% and the F1-score is 77.9%. The robustness of the proposed approach can be also appreciated from the ROC diagram [33] of Fig. 12. The AUC is about 85%. Positive

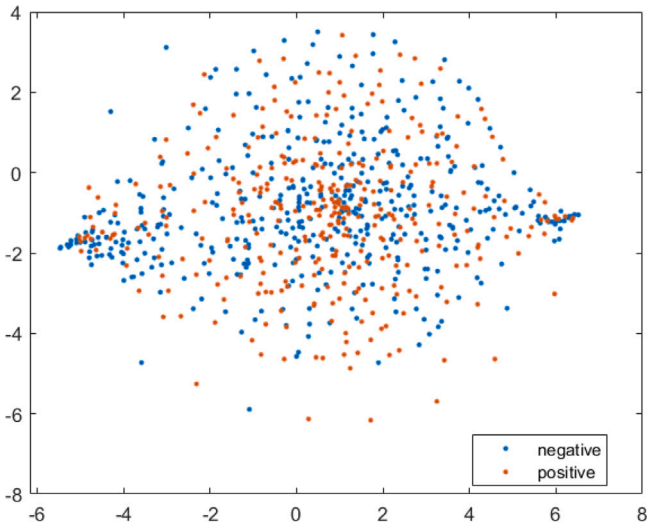


Fig. 9. t-SNE visualization of the original (acquired) data set.

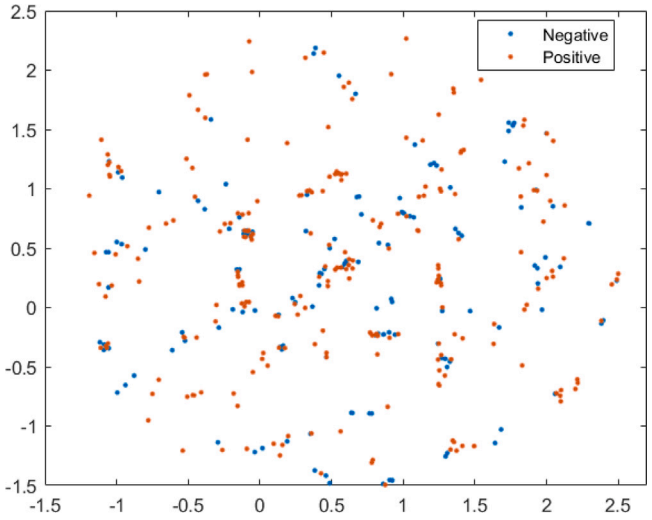


Fig. 10. t-SNE visualization of the harmonic components after HPSS.

and negative model operating points are evidenced for the sake of completeness. The explainability of the pipeline is investigated through the gradient-weighted class activation mapping (Grad-CAM) [34]. Figs. 13 and 14 show the Grad-CAM for the negative and positive STFT of Figs. 7 and 8, respectively. Considering the resizing of the STFT before feeding the CNN, we can infer that the Grad-CAM of the negative STFT is focused on the highest band of the spectrogram, i.e. 800–2000 Hz, where the absence of components with a relevant power can be interpreted as a lack of abnormal lung sounds. On the contrary, the Grad-CAM of the positive STFT denotes an attention to the middle band of the spectrogram, i.e. 300–1000 Hz, where the components of velcro crackles are significant. These results are consistent with those of previous works [19,23].

The pipeline shown in Fig. 1 has been compared with 3 distinct versions embodying well known components. The former is composed by high-pass filtering, VME and continuous wavelet transform (CWT). The second includes VME and gammatone filterbank based on 32 filters. The latter employs the ResNet-18 [35] as CNN in place of the Googlenet. All the solutions have been tested on the same data set described in Section 2. The ILDNet [11] has been also considered for comparison as state of the art. In fact, this framework has been designed to detect the abnormal respiratory sounds related to ILD. In

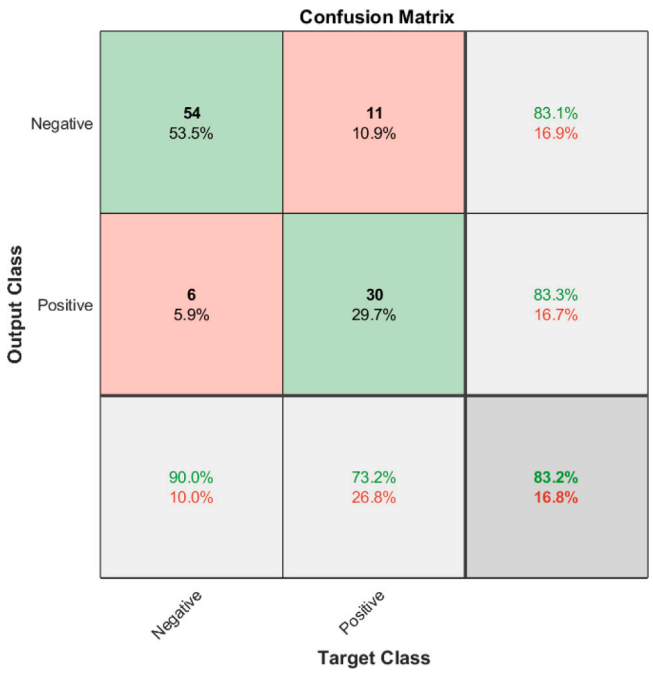


Fig. 11. Confusion matrix of the proposed pipeline for the classification of lung sounds over the test set.

Table 2
Accuracy, sensitivity, specificity, precision and F1-score of the proposed pipeline for the classification of lung sounds.

Metric	Result
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN} = 83.2\%$
Sensitivity	$\frac{TP}{TP+FN} = 73.2\%$
Specificity	$\frac{TN}{TN+FP} = 90.0\%$
Precision	$\frac{TP}{TP+FP} = 83.3\%\%$
F1-score	$\frac{2*Precision*Sensitivity}{Precision+Sensitivity} = 77.9\%$

Table 3
Performance of the proposed pipeline (with Googlenet) compared to 3 distinct variants. The ILDNet [11] has been considered as state of the art.

	Acc.	Sens.	Spec.	Prec.	F1-score
Proposed pipeline of Fig. 1 with Googlenet as CNN	83,2%	73.2%	90.0%	83.3%	77.9%
Proposed pipeline of Fig. 1 with ResNet-18 as CNN	83.2%	75,6%	88,3%	81,6%	78,5%
High-pass filter, VME, CWT, Googlenet	81,2%	68,3%	90,0%	82,4%	74,7%
VME, gammatone filterbank, Googlenet	81,2%	80,5%	81,7%	75,0%	77,7%
ILDNet [11]	81,3%	78,9%	83,3%	80,4%	79,6%

particular, the data set of [11] is composed by the respiratory sounds of 17 patients affected by ILD and of 8 healthy subjects, all collected in the BRACET [12] data set. Further healthy sounds are picked from the KAUH [13] to balance the overall data set. The performance of these pipelines are summarized in Table 3. The proposed solution outperforms its counterparts in terms of accuracy, specificity and precision. This approach suffers in terms of sensibility with respect to the gammatone time-frequency analysis, probably because the filterbank is able to evidence some features given bands. The robustness of our technique is witnessed by the F1-score similar to that of its counterparts. The influence of adopting the ResNet-18 for classification in place of the Googlenet is minimal.

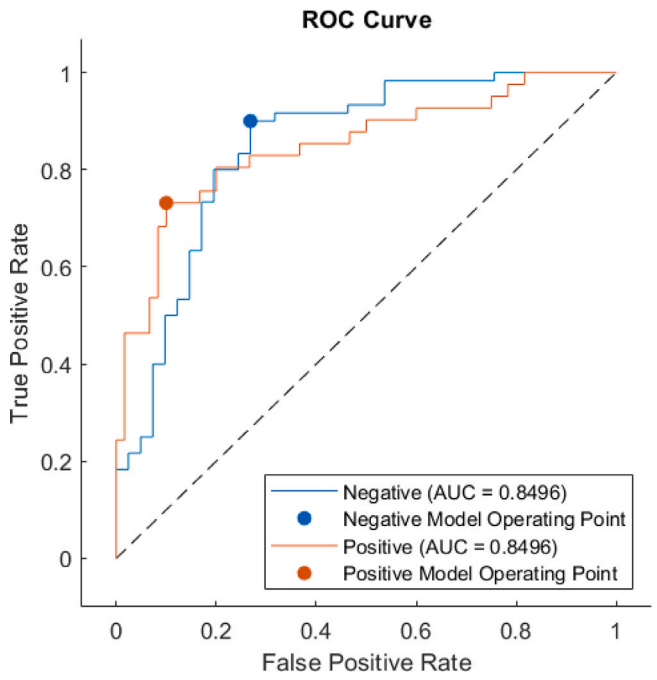


Fig. 12. ROC diagram and AUC of the test set.

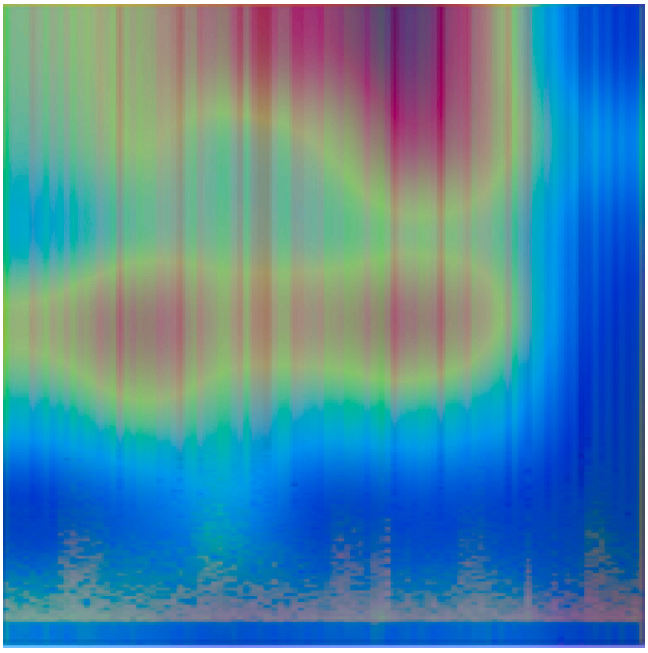


Fig. 13. Grad-CAM for the negative STFT.

From a clinical perspective, the ultimate scope of the proposed tool consists of raising the diagnostic suspicion of ILD for each patient, rather than for each auscultation. To this aim, the test set has been composed by 108 auscultations related to 18 patients, 9 negative and 9 positive to RA-ILD. The 18 patients of the test set are randomly picked out from the whole data set. The auscultations related to the remaining 119 patients are employed for training and validation. The classification of each patient in the test set is given by averaging the prediction probabilities of the 6 auscultations (per patient). This approach has been repeated for 5 distinct test sets, where each test set is perfectly

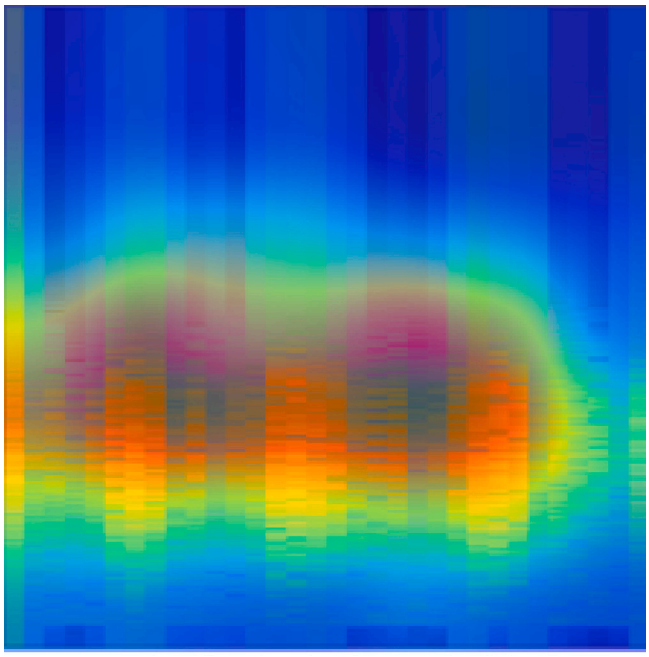


Fig. 14. Grad-CAM for the positive STFT.

Table 4
Accuracy, sensitivity, specificity, precision and F1-score of the proposed pipeline for the classification of patients. T. set stands for test set.

Metric	T. set 1	T. set 2	T. set 3	T. set 4	T. set 5	Average
Accuracy	83.3%	88.9%	83.3%	88.9%	94.4%	87.8%
Sensitivity	77, 8%	88.9%	77, 8%	77, 8%	88, 9%	82, 2%
Specificity	88, 9%	88.9%	88, 9%	100%	100%	93, 3%
Precision	87, 5%	88.9%	87, 5%	100%	100%	92, 5%
F1-score	82, 4%	88.9%	82, 4%	87, 5%	94, 1%	87, 1%

uncorrelated to the other ones, in the sense that each patient may appear in one test set only. The results of this approach are summarized in Table 4. We refer to the average performance, i.e. to the last column of Table 4, in the following to generalize the discussion as much as possible. Sensitivity, specificity and f1-score are all enhanced by the decision fusion on patients with respect to the corresponding metrics devised for single auscultations. Sensitivity, a very important feature for the considered clinical problem, is increased to 82,2%. Specificity is raised to 93,3%, the F1-score is largely improved to a competitive 87,1%. Another improvement entailed by combining the auscultation predictions of the same patient is precision, that is increased from 83,3% to 92,5%. This remarkable result can be motivated as follows. Auscultations misclassified into FP are mostly generated by artifacts or physiological reasons, like for instance cough and sputum. These noise sources usually do not affect more than 1 or 2 auscultations per patient. Then, TN predictions are capable to compensate for FP predictions within the lung sounds of the same patient. The significant improvement in the sensitivity results in an overall accuracy of 87.8% in the detection of ILD in RA patients. Finally, it is worth pointing out that the performance achieved over all the 5 test sets is similar to or better than that obtained in the classification of single samples of lung sounds (see Table 2). This means that the proposed pipeline is capable to handle the whole available data set, i.e. in other words there are no subset of “unmanageable” data.

5. Conclusions

This work presents a pipeline for pre-processing and classification of lung sounds aimed at the detection of ILD in patients affected by

RA. The pipeline is composed by high pass filtering and VME for denoising, sample cutting for data augmentation, STFT and HPSS for time-frequency analysis and enhancement of pathological lung sounds, CNN trained with transfer learning for binary (positive or negative to RA-ILD) classification. The proposed solution has been tested on a data set of lung sounds collected in a clinical study performed at the university hospital of Modena (Italy). The data set consists of 137 patient affected by RA, with 6 auscultations per patient. The ground truth is represented by the HRCT report. Our system evidenced an overall accuracy of 83.2% in the classification of lung sounds, with a F1-score of 77,9%. Combining the predictions provided by the CNN on the 6 auscultations per patients, the overall accuracy is increased to 87.8% with a F1-score of 87,1%. This performance makes the proposed solution a formidable candidate tool for screening ILD in patients affected by RA. In fact, physical lung auscultation followed by digital signal processing is: (a) safe for patients that are not exposed to ionizing radiation; (b) cheap for the national health system with respect to HRCT. Consequently, our approach has the potential to improve the follow-up and life expectation of RA-ILD patients, as well as the quality of life of RA patients.

CRediT authorship contribution statement

Fabrizio Pancaldi: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Luca Dibiase:** Writing – original draft, Visualization, Validation, Software, Investigation, Formal analysis, Data curation.

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.

Acknowledgment

This work has been partially supported by the research fund “Fondo di Ateneo per la Ricerca (FAR) 2024” allocated by the University of Modena and Reggio Emilia (Italy).

Data availability

The data that has been used is confidential.

References

[1] Y. Alamanos, A. Drosos, Epidemiology of adult rheumatoid arthritis, Autoimmun. Rev. 4 (3) (2005) 130–136, http://dx.doi.org/10.1016/j.autrev.2004.09.002.

[2] C. Hyldgaard, O. Hilberg, A.B. Pedersen, S.P. Ulrichsen, A. Løkke, E. Bendstrup, T. Ellingsen, A population-based cohort study of rheumatoid arthritis-associated interstitial lung disease: comorbidity and mortality, Ann. Rheum. Dis. 76 (10) (2017) 1700–1706, http://dx.doi.org/10.1136/annrheumdis-2017-211138.

[3] G. Koduri, S. Norton, A. Young, N. Cox, P. Davies, J. Devlin, J. Dixey, A. Gough, P. Prouse, J. Winfield, P. Williams, Interstitial lung disease has a poor prognosis in rheumatoid arthritis: results from an inception cohort, Rheumatol. 49 (8) (2010) 1483–1489, http://dx.doi.org/10.1093/rheumatology/keq035.

[4] G. Sgalla, S.L.F. Walsh, N. Sverzellati, S. Fletcher, S. Cerri, B. Dimitrov, D. Nikolic, A. Barney, F. Pancaldi, L. Larcher, F. Luppi, M.G. Jones, D. Davies, L. Richeldi, “Velcro-type” crackles predict specific radiologic features of fibrotic interstitial lung disease, BMC Pulm. Med. 18 (1) (2018) http://dx.doi.org/10.1186/s12890-018-0670-0.

[5] Z. Sun, ICBHI 2017 challenge, 2023, http://dx.doi.org/10.7910/DVN/HT6PKI.

[6] N. Baghel, V. Nangia, M.K. Dutta, ALSD-net: Automatic lung sounds diagnosis network from pulmonary signals, Neural Comput. Appl. 33 (24) (2021) 17103–17118, http://dx.doi.org/10.1007/s00521-021-06302-1.

[7] 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 self-developed open-access lung sound database – HF lung V1, PLoS One 16 (7) (2021) e0254134, http://dx.doi.org/10.1371/journal.pone.0254134.

[8] L. Shi, K. Du, C. Zhang, H. Ma, W. Yan, Lung sound recognition algorithm based on vggish-bigru, IEEE Access 7 (2019) 139438–139449, http://dx.doi.org/10.1109/access.2019.2943492.

[9] R. Phettom, N. Theera-Umpon, S. Auephanwiriyakul, Automatic identification of abnormal lung sounds using time-frequency analysis and convolutional neural network, in: 2023 15th International Conference on Information Technology and Electrical Engineering, ICITEE, IEEE, 2023, pp. 1–6, http://dx.doi.org/10.1109/icitee59582.2023.10317776.

[10] F. Majzoobi, M.B. Khodabakhshi, S. Jamasb, S. Goudarzi, ConvLSNet: A lightweight architecture based on ConvLSTM model for the classification of pulmonary conditions using multichannel lung sound recordings, Artif. Intell. Med. 154 (2024) 102922, http://dx.doi.org/10.1016/j.artmed.2024.102922.

[11] A. Roy, U. Satija, ILDNet: A novel deep learning framework for interstitial lung disease identification using respiratory sounds, in: 2024 International Conference on Signal Processing and Communications, SPCOM, IEEE, 2024, pp. 1–5, http://dx.doi.org/10.1109/spcom60851.2024.10631581.

[12] 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.d. Carvalho, R.P. Paiva, BRACETS: Bimodal repository of auscultation coupled with electrical impedance thoracic signals, Comput. Methods Programs Biomed. 240 (2023) 107720, http://dx.doi.org/10.1016/j.cmpb.2023.107720.

[13] M. Fraiwan, L. Fraiwan, B. Khassawneh, A. Ibnian, A dataset of lung sounds recorded from the chest wall using an electronic stethoscope, Data Brief 35 (2021) 106913, http://dx.doi.org/10.1016/j.dib.2021.106913.

[14] A. Roy, U. Satija, A novel melspectrogram snippet representation learning framework for severity detection of chronic obstructive pulmonary diseases, IEEE Trans. Instrum. Meas. 72 (2023) 1–11, http://dx.doi.org/10.1109/tim.2023.3256468.

[15] A. Roy, U. Satija, A novel multi-head self-organized operational neural network architecture for chronic obstructive pulmonary disease detection using lung sounds, IEEE/ACM Trans. Audio Speech Lang. Process. 32 (2024) 2566–2575, http://dx.doi.org/10.1109/taslp.2024.3393743.

[16] J.F. Gemmeke, D.P.W. Ellis, D. Freedman, A. Jansen, W. Lawrence, R.C. Moore, M. Plakal, M. Ritter, Audio set: An ontology and human-labeled dataset for audio events, in: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2017, http://dx.doi.org/10.1109/icassp.2017.7952261.

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

[18] A. Roy, U. Satija, Effect of auscultation hindering noises on detection of adventitious respiratory sounds using pretrained audio neural nets: A comprehensive study, IEEE Trans. Instrum. Meas. 74 (2025) 1–8, http://dx.doi.org/10.1109/tim.2025.3571143.

[19] F. Pancaldi, M. Sebastiani, G. Cassone, F. Luppi, S. Cerri, G. Della 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.

[20] A. Manfredi, G. Cassone, S. Cerri, V. Venerito, A.L. Fedele, M. Trevisani, F. Furini, O. Addimanda, F. Pancaldi, G. Della 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.

[21] A. Manfredi, G. Cassone, C. Vacchi, F. Pancaldi, G. Della Casa, S. Cerri, L. De 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.

[22] 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.

[23] 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.

[24] 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.

[25] A. Fava, B. Dianat, A. Bertacchini, A. Manfredi, M. Sebastiani, M. Modena, F. Pancaldi, Pre-processing techniques to enhance the classification of lung sounds based on deep learning, Biomed. Signal Process. Control. 92 (2024) 106009, http://dx.doi.org/10.1016/j.bspc.2024.106009.

[26] J. Driedger, M. Muller, S. Ewert, Improving time-scale modification of music signals using harmonic-percussive separation, IEEE Signal Process. Lett. 21 (1) (2014) 105–109, http://dx.doi.org/10.1109/lsp.2013.2294023.

[27] 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.

[28] F.C. Arnett, S.M. Edworthy, D.A. Bloch, D.J. Mcshane, J.F. Fries, N.S. Cooper, L.A. Healey, S.R. Kaplan, M.H. Liang, H.S. Luthra, T.A. Medsger, D.M. Mitchell, D.H. Neustadt, R.S. Pinals, J.G. Schaller, J.T. Sharp, R.L. Wilder, G.G. Hunder, The American rheumatism association 1987 revised criteria for the classification of rheumatoid arthritis, Arthritis Rheum. 31 (3) (1988) 315–324, http://dx.doi.org/10.1002/art.1780310302.

[29] D. Aletaha, T. Neogi, A.J. Silman, J. Funovits, D.T. Felson, C.O. Bingham, N.S. Birnbaum, G.R. Burmester, V.P. Bykerk, M.D. Cohen, B. Combe, K.H. Costenbader, M. Dougados, P. Emery, G. Ferraccioli, J.M.W. Hazes, K. Hobbs, T.W.J. Huizinga, A. Kavanaugh, J. Kay, T.K. Kvien, T. Laing, P. Mease, H.A. Ménard, L.W. Moreland, R.L. Naden, T. Pincus, J.S. Smolen, E. Stanislawska-Biernat, D. Symmons, P.P. Tak, K.S. Upchurch, J. Vencovsky, F. Wolfe, G. Hawker, 2010 rheumatoid arthritis classification criteria: An American college of rheumatology/European league against rheumatism collaborative initiative, Arthritis Rheum. 62 (9) (2010) 2569–2581, http://dx.doi.org/10.1002/art.27584.

[30] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, 2015, pp. 1–9, http://dx.doi.org/10.1109/cvpr.2015.7298594.

[31] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, ImageNet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, http://dx.doi.org/10.1109/cvpr.2009.5206848.

[32] L. van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (86) (2008) 2579–2605, URL http://jmlr.org/papers/v9/vandermaaten08a.html.

[33] A.P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern Recognit. 30 (7) (1997) 1145–1159, http://dx.doi.org/10.1016/s0031-3203(96)00142-2.

[34] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual explanations from deep networks via gradient-based localization, in: 2017 IEEE International Conference on Computer Vision, ICCV, IEEE, 2017, pp. 618–626, http://dx.doi.org/10.1109/iccv.2017.74.

[35] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, 2016, pp. 770–778, http://dx.doi.org/10.1109/cvpr.2016.90.