Biomedical Signal Processing and Control 82 (2023) 104555

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Cochleogram-based adventitious sounds classification using convolutional neural networks☆

L.D. Mang [a,*], F.J. Canadas-Quesada [a], J.J. Carabias-Orti [a], E.F. Combarro [b], J. Ranilla [b]

[a] Departament of Telecommunication Engineering. University of Jaen, Campus Cientifico-Tecnologico de Linares, Avda. de la Universidad, s/n, Linares (Jaen), 23700, Spain

[b] Department of Computer Science, University of Oviedo, Campus de Gijón s/n, Gijon (Asturias), 33203, Spain

ARTICLE INFO

Dataset link: https://bhichallenge.med.auth.gr/ICBHI_2017_Challenge

Keywords:
Classification
Adventitious sounds
Wheezes
Crackles
STFT spectrogram
Mel-scaled spectrogram
Cochleogram
Convolutional neural network
Accuracy

ABSTRACT

Background: The World Health Organization (WHO) establishes as a top priority the early detection of respiratory diseases. This detection could be performed by means of recognizing the presence of acoustic biomarkers (adventitious sounds) from auscultation because it is still the main technique applied in any health center to assess the status of the respiratory system due to its non-invasive, low-cost, easy to apply, fast to diagnose and safe nature.

Method: Despite the novel deep learning approaches applied in this biomedical field, there is a notable lack of research that rigorously focuses on different time–frequency representations to determine the most suitable transformation to feed data into Convolutional Neural Network (CNN) architectures. In this paper, we propose the use of the cochleogram, based on modeling the frequency selectivity of the human cochlea, as an improved time–frequency representation to optimize the learning process of a CNN model in the classification of respiratory adventitious sounds. Our proposal is evaluated using the largest and most challenging public database of respiratory sounds.

Results: The cochleogram obtains the best binary classification results among the compared methods with an average accuracy of 85.1% in wheezes and 73.8% in crackles, and a competitive performance evaluating a multiclass classification scenario in comparison with other well-known state-of-the-art deep learning models.

Conclusion: The cochleogram provides a suitable time–frequency representation since it is able to model respiratory adventitious content more accurately by means of non-uniform spectral resolution and due to its increased robustness to noise and acoustic changes. This fact implies a significant improvement in the learning process of CNN models applied in the classification of respiratory adventitious sounds.

1. Introduction

The World Health Organization (WHO) warns about the importance of respiratory health and considers it a top priority in global decision-making, identifying the main respiratory diseases due to their severity and number of deaths worldwide [1]: (a) Chronic Obstructive Pulmonary Disease (COPD) affects more than 200 million people, and is underdiagnosed by 72 to 93%. Specifically, the direct cost of COPD accounts for 6% of total healthcare expenditure (€38,6 billion per year) in the European Union and represents 56% of the total cost of treatment of respiratory diseases [2]; (b) Asthma affects up to 334 million people and it is often not detected early, causing about 489,000 deaths annually [3]; (c) Lower respiratory tract infections (LRTIs) and pneumonia are two of the leading causes of death, with pneumonia being the leading cause of death in children under 5 years of age [4]; (d) Tuberculosis (TB) was suffered by almost 10,4 million people in 2015 [5]; and (e) Lung Cancer (LC) causes about 1,6 million deaths, accounting for 19,4% of the total deaths attributable to cancer in 2012 [6]. Nevertheless, WHO also alerts that there are other respiratory disorders such as sleep breathing (or sleep apnea), pulmonary hypertension and pulmonary embolism that are gaining importance as they affect an increasing number of the population in the last decades.

☆ This work was supported in part under grant PID2020-119082RB-C21,C22 funded MCIN/AEI/10.13039/501100011033, grant 1257914 funded by Programa Operativo FEDER Andalucia 2014–2020, grant P18-RT-1994 funded by the Ministry of Economy, Knowledge and University, Junta de Andalucía, Spain, grant AYUD/2021/50994 funded by Gobierno del Principado de Asturias, Spain and QUANTUM SPAIN project funded by the Ministry of Economic Affairs and Digital Transformation of the Spanish Government and the European Union through the Recovery, Transformation and Resilience Plan - NextGenerationEU.

* Corresponding author.
E-mail address: lmang@ujaen.es (L.D. Mang).

https://doi.org/10.1016/j.bspc.2022.104555
Received 27 August 2022; Received in revised form 13 November 2022; Accepted 26 December 2022
Available online 4 January 2023

Although auscultation requires acoustically trained physicians to correctly detect and recognize adventitious sounds, auscultation is still the earliest diagnosis in any health center facility to assess the status of the respiratory system as it is non-invasive, low-cost, easy to apply, fast to diagnose and safe [7] but further medical tests may be performed to confirm or refine the diagnosis [8]. Thus, when a physician performs auscultation to analyze the state of health of a patient's respiratory system, he/she usually hears adventitious respiratory sounds (ARS) superimposed with the normal respiratory sounds, since the presence of adventitious sounds implies the existence of pulmonary pathologies, which allows early detection of respiratory disorders in order to prevent the patient from not receiving the appropriate medical treatment and, consequently, from returning to the hospital with an aggravation of the initially undetected disorder. As a result, a great effort, both from the medical and the signal processing and artificial intelligence community, is being applied to the early detection of these acoustic bio-markers (adventitious sounds) to be considered of utmost relevance in the overall health diagnosis as early detection implies both a more effective medical treatment of the patient as well as a better management of economic resources as it is only a fraction of the cost of treatment in a patient not diagnosed early [9].

In a general categorization, respiratory sounds are classified into normal and abnormal (adventitious sounds) according to the Computed Respiratory Sound Analysis (CORSA) [10]. Normal respiratory sounds (RS) appear in healthy lungs and contain a broadband spectrum that locates most of the energy between 60–1000 Hz [11]. Although there are many types of adventitious sounds superimposed on RS (for example, wheezes, crackles, pleural rub, stridor, and squawks), the main ones are considered to be wheezes and crackles. Wheezes (WS) are continuous and musical sounds that usually present a pitch located between 100–1000 Hz with a duration greater than 100 ms, showing narrowband spectral trajectories commonly known as "snakes" due to the temporal evolution of the energy in the frequency domain [7,12]. Crackles (CS) are discontinuous, short and explosive sounds whose spectral energy is located between 100–2000 Hz [13]. Specifically, crackles are classified into coarse and fine. Coarse crackles have a temporal duration below than 20 ms and usually have a low pitch around 350 Hz due to the presence of air bubbles in the large bronchi. In contrast, fine crackles usually have a maximum temporal duration of smaller than 5 ms with a high pitch located around 650 Hz, usually caused by the sudden opening of the small airways [14]. Pathologically, wheezing is associated with obstructive lung diseases such as COPD, asthma, bronchiolitis, bronchitis, bronchiectasis, or emphysema [15,16]. Instead, crackles are associated with specific lung diseases such as pneumonia, interstitial pulmonary fibrosis, pulmonary edema, or idiopathic pulmonary fibrosis.

In the last decades, many works have been proposed in the analysis of the most common adventitious respiratory sounds, specifically, crackles and wheezes. Focusing on the detection and classification of crackles sounds, most of the approaches are based on spectrogram analysis [17,18], auto-regressive (AR) models [19,20] wavelet transform [21–24], fractal dimension filtering [25–28], entropy [29,30], empirical mode decomposition (EMD) [31], fuzzy systems [32], gaussian mixture models (GMM) [33], logistic regression [34], support vector machines (SVM) [35–37], independent component analysis (ICA) [38], multi perceptron networks (MPN) [39], non-negative matrix factorization (NMF) [40], convolutional neural networks (CNN) [41,42], recurrent neural networks (RNN) [43] and hybrid neural networks [44]. Focusing on the detection and classification of wheezing sounds, most of the approaches are based on spectrogram analysis [45–49], Higher-order statistics [50], AR model [51], auditory modeling [52], wavelet transform [53–55], tonal index [56,57], GMM [58,59], entropy [60], Mel-Frequency Cepstral Coefficients (MFCCs) [61], Neural Networks (NN) [62,63], Hidden Markov Model (HMM) [64] and NMF [65,66].

However, automatic classification of respiratory sounds has been hindered by the lack of a large set of clinical respiratory data, until the appearance of the public largest database ICBHI in 2019 [67,68], since the process of acquiring and labeling respiratory sounds is very laborious by consuming a lot of time and human resources. As a consequence, several methods based on signal processing [69–71] were developed to classify four types of respiratory sounds (such as, normal, crackles, wheezes and both crackles and wheezes) using the ICBHI database. The emergence of a large number of novel works based on convolutional neural network (CNNs) architectures applied in the biomedical field of adventitious respiratory sound analysis [72–80,80–98] obtained promising performances as had already occurred in other areas, such as audio [99,100], image [101,102] or biomedicine [103,104], since CNNs reduce the human error made when applying feature extraction techniques in conventional methods. This error reduction is due to the fact that CNNs are able to automatically learn the most relevant temporal and spectral features shown by respiratory sounds from large datasets [105]. In [72], authors indicated that classifying spectrogram images with CNN provides similar results compared to support vector machine (SVM), and given the large amount of data, CNN and SVM can accurately classify and pre-diagnose respiratory audio. Bardou et al. [41] used CNNs using spectrograms, MFCC and local binary pattern (LBP) features. Results reported that CNNs can replace conventional classifiers through the use of fully-connected layers to train the features (MFCC and LBP), summing up the output of Softmax activation of four CNN models. In [106], a scalogram based optimized AlexNet pre-trained CNN model is developed to extract the visual details from the pixel values of lung sound images and for accurate classification and detection. Demir et al. [80], by means of Short Time Fourier Transform (STFT), proposed two deep learning based approaches for lung sound classification. In the first approach, a pre-trained deep CNN was used for feature extraction and a SVM classifier was used in classification of the lung sounds. In the second approach, the pre-trained deep CNN model was fine-tuned (transfer learning) via spectrogram images for lung sound classification. The classification performance were tested by using the ten-fold cross validation, obtaining accuracies for the first and second proposed methods equals 65.5% and 63.09%. Zulfiqar et al. [87] proposed a Fourier approach in order to classify seven abnormal respiratory sounds based on the spectrum analysis using Artificial Noise Addition (ANA) in conjunction with CNN architectures. The purpose of ANA was to add the artificial noise of exact spectral nature to enhance the actual spectrogram of faded sounds to make them more strengthen and enhance their robustness. This robustness, associated to sound features to be identified more accurately than the respiratory sounds without ANA, increased the classification performance of the proposal. Finally, a set of algorithms were evaluated using different spectrograms and the optimal results were obtained using the AlexNet algorithm. Chanane and Bahoura [86] proposed a CNN architecture to improve the classification of respiratory sounds recorded by electronic stethoscopes analyzing the impact of variant frequency representations techniques using data normalization on top of data augmentation techniques. In [107], authors demonstrated the ability of CNN and bidirectional long short-term memory units (BiLSTM) to recognize pulmonary diseases from lung sounds in order to extract time-domain features. Results indicated an overall average accuracy of 99.62%. Tariq et al. [93] proposed a CNN model-based fusion (FDC) from the three feature-based CNNs models to classify lung and heart disease. It reported that is more effective to classify heart or lung diseases with images transformed from three different sound features, such as Spectrogram, MFCC, and Chromagram. Finally, three types of data augmentation, such as Noise, Pitch-Shift, and Time-Stretch, have been effectively applied for optimal training and testing. Nguyen and Pernkopf [95] exploited transfer learning applied to architectures of residual neural networks. Specifically, Batch Normalization is replaced in order to avoid poor performance in case of a data distribution shift between training and test data. Stochastic normalization is applied in each residual block of the pre-trained architecture to reduce the effect

of over-fitting on small datasets. Petmezas et al. [98] developed a CNN–LSTM model using the focal loss function with training data imbalance. Respiratory cycles were converted into STFT time–frequency representation images and processed with CNN in order to extract their most predominant features. These features were used with LSTM identifying long-term dependencies between them. In Rocha et al. [108], an exhaustive evaluation is presented analyzing combinations of several adventitious sounds by means of linear discriminant analysis (LDA), SVM, boosted trees (RUSBoost) and CNNs using ICBHI [68]. The LDA, SVM and RUSBoost classifiers were fed features extracted from the spectrograms, including some novel acoustic features. On the other hand, the CNNs received STFT and Mel spectrograms as inputs. Authors reported that while CNNs have become state-of-the-art solutions in several tasks, they were not enough to tackle this problem emphasizing that there is still room for improvement in CNN-based respiratory sound classification focusing on alternative time–frequency (TF) representations. In this paper, we focus on CNN-based approaches applied to respiratory sounds classification.

Despite of the advanced machine learning techniques used in the aforementioned works, there is a notable lack of research that rigorously evaluates the different TF representations to determine the most appropriate transformation to feed data into the systems since it is well known that remarkable differences can be observed using different TF representations to replicate the performance of the human ear [109,110]. In this work, we study the effect of the classical TF representations (STFT and Mel spectrogram) and propose the use of a human auditory based non-linear representation called cochleogram which has already been applied in other scientific areas such as audio [111,112] or heart sound detection [113] but, to our best knowledge, not to classify adventitious respiratory sounds. In particular, the non-uniformity of this representations has demonstrated higher robustness against noise and acoustic changes than the classical linear or speech-based time–frequency signal representations [113].

In order to demonstrate the benefits of using the human auditory system based cochleogram rather than the standard time–frequency representations (i.e. STFT and Mel-scaled spectrogram), we propose a CNN architecture similar to the one presented in [108] and evaluate the use of the different input time–frequency representation for the task of adventitious respiratory sounds classification using the largest and most challenging public database of breath sounds (ICBHI) [67,68]. In particular, we study the performance of our baseline CNN model to detect the occurrence of crackles and wheezing in a binary and a multiclass classification scenario. Finally, we study the effect of the studied time–frequency representations on other state-of-the-art CNN models including AlexNet [114], ResNet50 [115] and VGG16 [116].

The paper is organized as follows: the dataset, TF representations and the baseline CNN architecture used in this work are detailed in Section 2. The evaluation is described in Section 3 including the metrics, experimental setup and the parameter optimization. In Section 4, the classification results are presented and the performance of the baseline and other state-of-the-art CNN approaches compared using as input data the STFT spectrogram, Mel-scaled spectrogram and the cochleogram of individual respiratory cycles. Section 5 presents a discussion about the comparison between the proposed method and state-of-the art methods evaluating the four-classes (normal, wheezes, crackles and both (crackles+wheezes)) classification performance in the ICBHI database. Finally, conclusions and future work are addressed in Section 6.

2. Materials and methods

This section details the characteristics of the dataset evaluated as well as the formal concepts for each TF representation and finally, the employed network architecture.

Table 1
Cycle breakdown of ICBHI 2017 challenge dataset.

Number of cycles	Total
Crackles	1.864
Wheezes	886
Crackles + Wheezes	506
Normal	3.642
Total number of cycles	**6.898**

2.1. Database

The publicly available ICBHI 2017 Challenge dataset [68] has been used in this work, which consists of 920 annotated recordings with lengths varying from 10–90 s with a total of 5.5 h. The sounds were recorded using three different digital stethoscopes, with sampling frequencies of 4 KHz, 10 KHz and 44.1 KHz, respectively. As depicted in Table 1, the classes of this dataset consist of crackles, wheezes, normal, and wheezes plus crackles. A sound file may include one or more class tag since the sound files are separated into respiratory cycles. Manual annotation about the start and end times for every class is also provided. Consequently, we divided the entire dataset into respiratory cycles using the text files appended with each audio file included in the database. In the ICBHI dataset, the length of breathing cycles ranges from 0.2 s to 16.2 s with a mean cycle length of 2.7 s. Although training CNN-based networks is possible using adaptive average pooling, this strategy uses to perform poorly in comparison with fixed size signals [91]. In this paper, a zero-padding was added to each respiratory cycle until a total length of 6 s was reached. Moreover, as the respiratory events of interest (i.e. wheezing and crackles) do not exceed 2 KHz [7,10,12], to standardize, we decided to downsample each respiratory cycle to $f_s = 4$ KHz.

2.2. Time–frequency representations

In this subsection, the mathematical and signal processing background associated to STFT spectrogram, Mel-scaled spectrogram and cochleogram is briefly described.

2.2.1. STFT spectrogram

The Short Time Fourier Transform (STFT) is the most widely used TF representation to perform signal analysis in the frequency domain. Given an input signal $x(t)$, which is converted to $x[n]$ using a sampling rate f_s in Hz, the STFT spectrogram $\mathbf{X}_c \in \mathbb{C}^{K\times L}$, composed of K frequency bins and L frames, calculates each coefficient $X_c(k,l)$ for each kth frequency bin and lth time frame as follows,

$$X_c(k,l) = \sum_{n=0}^{N-1} x\left[(l-1)\cdot J + n\right] w[n] e^{-j\frac{2\pi}{N}kn}, \tag{1}$$

where $w[n]$ is a N samples analysis window, J represents the samples time shift, $k = [0,\ldots,K-1]$ and $l = [1,\ldots,L]$. In general, classical approaches discard the phase information and the analysis of the spectral content is performed based only on the magnitude spectrogram $\mathbf{X}$, that is, $\mathbf{X} = |\mathbf{X}_c| \in \mathbb{R}_+^{K\times L}$.

However, STFT spectrograms may not be the optimal TF representation for analyzing respiratory sounds because STFT provides constant bandwidth, which implies lower resolution at low frequencies where most of the relevant respiratory spectral content is present. Moreover, it is well known that although STFT obtains satisfactory results in quiet conditions, STFT significantly reduces its performance analyzing auscultated sounds in noisy environments [113].

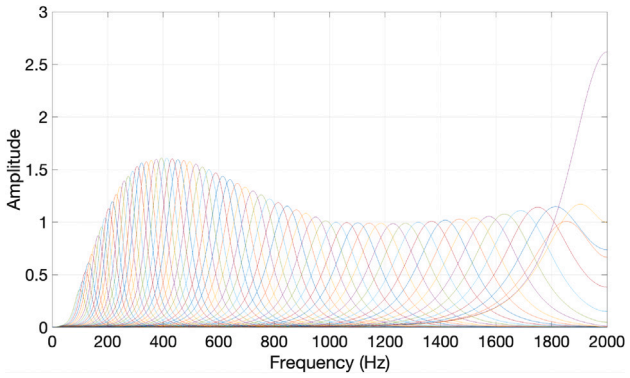

Fig. 1. Middle-ear gain normalization of the frequency response of 64-channel gammatone filter bank [111]. It can observed higher spectral resolution at low frequencies.

2.2.2. Mel-scaled spectrogram

The Mel scale is inspired by the human auditory system and physiological findings on speech perception [117]. In fact, the human ear is more sensitive to differences between lower frequencies than higher frequencies as well as the loudness is heard on a logarithmic scale rather than linear as shown in Eq. (2),

$$M(f) = 1127 \cdot log\left(1 + \frac{f(Hz)}{700}\right) \tag{2}$$

The Mel-scaled spectrogram is computed using the energies obtained from the so called Mel filter bank [117]. Specifically, the output of the cth filter is obtained using Eq. (3) as,

$$E(c,l) = log_{10}\left(\sum_{k=0}^{\frac{N}{2}-1} V(c,k) \cdot |X_c(k,l)|\right),$$
$$c = 1, 2, \ldots, C \tag{3}$$

where $E(c,l)$ is the energy associated to the cth filter applied in the lth frame, $V(c,k)$ is the normalized filter bank response which is equally spaced on the Mel scale [118], C is the total number of Mel filters, the kth bin of frequency associated to the STFT spectrogram $\mathbf{X}_c$ and N samples window. In this work, the number of Mel filters C has been set to 64 as occurs in [119,120].

This Mel-log or variants such as Constant-Q transform have been extensively used in the literature of speech and music signal processing [121]. In fact, this representation has shown moderate success in low noisy situations. However, adventitious sounds are superimposed on normal respiratory sounds and sometimes also on high noises from stethoscope rubbing together with different types of ambient noises. Therefore, more robust representations are needed to improve the performance against noise and environmental acoustic changes.

2.2.3. Cochleogram

The gammatone filter attempts to model the frequency selectivity of the human cochlea [122,123], using non-uniform spectral resolution by associating wider frequency bandwidths with higher frequencies in order to mimic the performance of the human ear as shown in Fig. 1. This variable resolution provides a TF representation capable of extracting more accurate spectral content from the input signal due to higher robustness against noise and acoustic changes [111,113,124].

The cochleogram is computed using a gammatone filter bank in which the impulse response of gammatone filter $g(t)$ is obtained multiplying a gamma distribution and a sinusoidal function [111,113],

$$g(t) = t^{o-1} e^{2\pi b(f_c)t} cos(2\pi f_c t), \qquad t > 0 \tag{4}$$

Table 2
Baseline CNN Architecture. .

Layer type	Kernel Atribute	Activation
Conv2D	3 × 3 32 Filters	LeakyReLu
MaxPool2D	2 × 2	–
Dropout	0.25	–
Conv2D	3 × 3 64 Filters	LeakyReLu
MaxPool2D	2 × 2	–
Dropout	0.25	–
Flatten	–	–
Dense	200 units	LeakyReLu
Dropout	0.5	–
Dense	200 units	Softmax/Sigmoid*

where the parameters are the filter order o and the exponential decay coefficient $b(f_c)$ associated to the center frequency f_c Hz that decides the bandwidth [111]. The center frequencies are equally spaced on the equivalent rectangular bandwidth (ERB) scale as shown in Eq. (6). In this work, we have selected the lower and upper central frequencies f_c equal to 100 Hz and $\frac{f_s}{2}$ Hz on the linear frequency scale because most adventitious respiratory sounds, mainly wheezing and crackles as previously mentioned in Section 1, contain most content in this spectral range. The order has been set $o = 4$ because it provides satisfactory results replicating the human auditory filter as depicted in [113].

$$b(f_c) = 1,019 \cdot ERB(f_c) \tag{5}$$

$$ERB(f_c) = 24,7 \cdot \left(4,37 \cdot \frac{f_c}{1000} + 1\right) \tag{6}$$

Then, the input signal $x[n]$ is filtered by each response filter $g(t)$ and subsequently, each filtered signal is windowed into frames of N samples with an offset equal to J samples in order to calculate the cochleogram by computing the power of each frame in each channel as occurs in [111].

Fig. 2 shows the TF representations computed by STFT, Mel-scaled and cochleogram. It can be observed that the adventitious sounds shown by the cochleogram are easier to identify and recognize compared to the other spectrograms. It seems that the low spectral respiratory content is more accurate modeled by the gammatone filtering by means of non-uniform resolution, a fact that does not occur in the STFT spectrogram and in a worse way in the Mel spectrogram that both of them tend to show more dispersed spectrograms in the frequency range in which respiratory sounds are often located [124]. Moreover, the cochleogram provides less uncertainty to define the spectro-temporal pattern associated with the wheeze found around 2.35 s compared to the other TF representations.

2.3. Neural network architecture

As our main objective in this work is performing a comparison between different TF representations, we decided to use as our baseline model a standard Convolutional Neural Network (CNN) model such as the one implemented in [108] as shown in Fig. 3. More details can be found in [108].

In order to perform the different experiments, we evaluated each TF representation individually as an input for the baseline CNN. Apart of the parameters shown in Fig. 3, a detailed table of the different layers is displayed in Table 2. A total of 30 epochs were used with a batch size of 16, a learning rate of 0.001 and the adaptive data momentum (ADAM) optimization algorithm. In order to avoid an over-fitting, the early stopping strategy employed during the training was set at 10 consecutive epoch, taking as monitor parameter the validation loss.

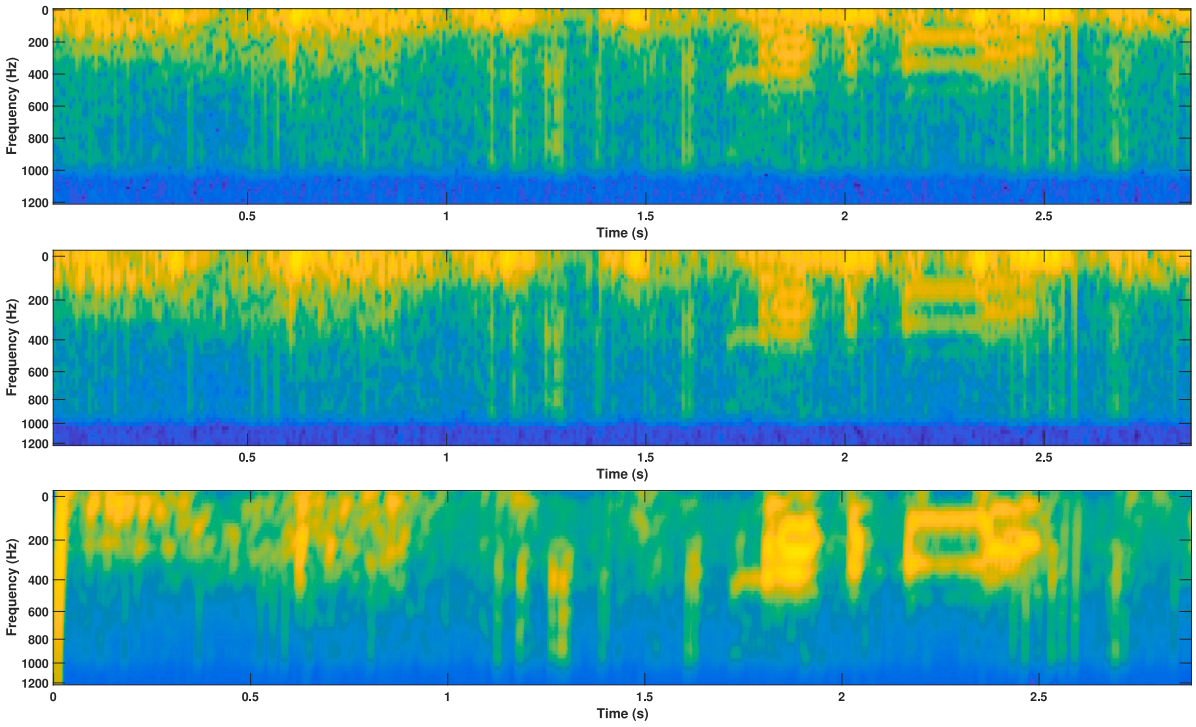

Fig. 2. Magnitude, in logarithmic scale, of the TF representations analyzing a respiratory cycle with a time duration of 2.9 s associated to the patient number 103 from ICBHI [67]. The respiratory cycle is composed by one wheeze sound located in the temporal range [2.1–2.6] s. STFT spectrogram (top), Mel-scaled spectrogram (center) and Cochleogram (bottom).

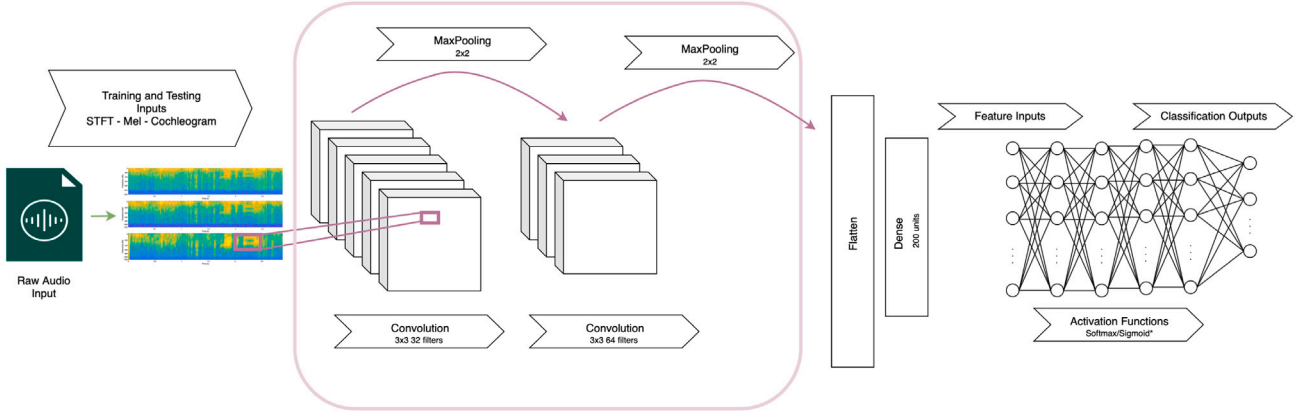

Fig. 3. Baseline CNN Network.

3. Evaluation setup

A systematic comparison is presented throughout this section analyzing the classification results provided by each TF representation using a set of CNN architectures. First, the different setups for the three TF representations will be presented. Then, the metrics employed to evaluate each experiment are described in Section 3.1. The training/testing conditions are detailed in Section 3.2. Finally, results for the tasks of detecting wheezing and crackles and the performance compared with other state-of-the-art methods are reported.

3.1. Metrics

To evaluate the performance of the developed model, several metrics were included in this study to analyze the classification confusion matrix: Accuracy (Acc), Sensitivity (Sen), Specificity (Spe), Precision ($Prec$) and Score (Sco). The confusion matrix was generated sequentially after every fold, and all evaluation metrics were calculated from the overall confusion matrix after the 10-fold cross validation of the training/classification scheme. For the purpose of these metrics, TP (true positive) indicates the number of correctly detected adventitious events, TN (true negative) is the number of correctly classified normal events, FP (False Positives) are events that are incorrectly classified as the adventitious class and FN (False Negatives) are events of the adventitious class that are incorrectly classified as normal events. These metrics are mathematically defined as:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}, \tag{7}$$

$$Prec = \frac{TP}{TP + FP}, \tag{8}$$

$$Sen = \frac{TP}{TP + FN}, \tag{9}$$

$$Spe = \frac{TN}{TN + FP}, \tag{10}$$

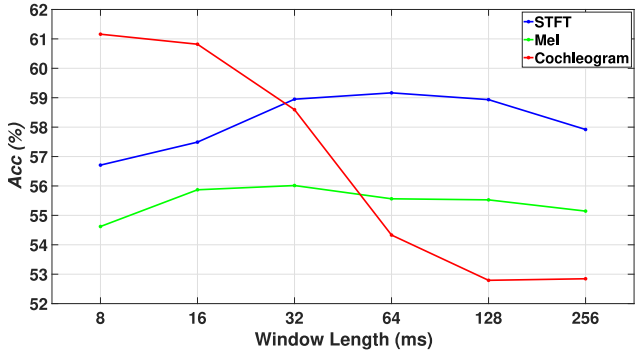

Fig. 4. Overall accuracy results evaluating 4 classes, in terms of mean values for the whole range of window lengths, using the ICBHI dataset.

$$Sco = \frac{Sen + Spe}{2}, \tag{11}$$

As occurs in [108,113], we selected the accuracy (Eq. (7) as the main general metric to study the effect of the parameters tuning whereas all the metrics are presented for comparison with other state-of-the-art methods in the standard ICBHI multiclass classification challenge [87,108].

3.2. Experimental setup

First, preliminary analysis was accomplished by evaluating several window lengths for each TF representation. Specifically, the following window lengths were tested, N = [8, 16, 32, 64, 128, 256] ms. Regarding the window type and the time shift between windows, we used a Blackman–Harris window and a 75% overlap size as this was the optimal setup in [108].

Regarding the training/testing conditions, in this paper, 10-fold cross validation was computed five times and the average results are reported [125]. For each fold, the dataset was divided using a 75%-25% distribution for the training–testing subsets respectively. Once the subsets were defined, a 25 percent from the training set was used for validation. Due to there is a unbalanced number of crackles and wheezes events as previously detailed in Table 1, we ensure that all of them have been distributed in an proportional manner in every fold.

The experimental works were applied using Tensorflow and Keras installed on a computer with an Intel(R) Core(TM) i7-5500 CPU @2.4 GHz with 4 core, NVIDIA GeForce GTX1080Ti GPU and 64 GB RAM. For reproducible research, the code can be publicly accessed at[1].

3.3. Optimal parameters estimation

As previously explained in Section 3.2, different window lengths are studied for each TF representation. In order to estimate the optimal parameters, we evaluate the classification performance between four classes: normal, crackles, wheezes and both (crackles + wheezes). Fig. 4 shows the averaged accuracy values obtained from the 10-fold cross-validation for each TF representation used as input for our baseline CNN model. As can be seen in Fig. 4, the best performance with the STFT-based model is obtained between 32 and 128 ms and the optimal performance is obtained using a 64 ms window length. Regarding the Mel-scaled model, the best performance is obtained with a 32 ms window length, these optimal values for both the STFT and Mel models being in line with those presented in [108].

On the contrary, in the case of the human auditory system based cochleogram, better performance was obtained using smaller window sizes which confirms previous studies performed in [124]. In fact, the optimal value was obtained using 8 ms window length as shown in Fig. 4. This optimal window length seems to be more suitable for modeling the explosive and transient nature that crackle sounds often exhibit. Hence, worse performance is obtained as the window length increases, as the high temporal resolution required for the correct modeling of such short time signals is lost, causing the system to become confused at a higher rate. Note that, as explained in Section 2.2.3, the human auditory based filterbank used in this transform operates in the logarithmic ERB scale, however, a windowing is required for each resulting channel of the filtering process in order to obtain a discrete TF representation. Initially, preliminary analysis based on the Multi-Resolution cochleogram (MRCG) approach [111] were analyzed in which a high temporal resolution cochleogram and several low temporal resolution cochleograms were combined attempting to model local and temporal context information. Results indicated that using only a high temporal resolution cochleogram provides better classification performance for respiratory adventitious sounds compared to the combination of the previous cochleograms of different temporal resolution. This fact suggests that a high temporal resolution is a crucial feature to be taken into account in order to improve the modeling of crackle sounds. For this reason, only the TF representation based on a single high temporal resolution cochleogram has been used throughout the paper.

4. Classification results

Further experiments are conducted to assess the classification performance of the proposed TF representation once the parameter optimization has been addressed.

4.1. Binary classification results

Here, the classification performance of the cochleogram is evaluated on a binary classification problem where the objective is to detect the occurrence of crackles or wheezes when the input is composed of a monaural sound signal corresponding to individual respiratory cycles.

Results of the classification performance of the baseline model using as input the three TF representations (STFT, Mel-scaled spectrogram and the Cochleogram) are shown in Fig. 5. It can be observed that the best performance, for both crackles and wheezing classification, is provided by the Cochleogram representation (an average Acc = 85.1% for wheezes and Acc = 73.8% for crackles). It is interesting to observe that the STFT representation provides competitive results (Acc = 84.9% for wheezes and Acc = 71.5% for crackles) and clearly outperforms the Mel-scaled representation. In fact, applying a low-pass filtering to the frequencies of interest and using a proper window length and hop size provides sufficient resolution to accurate detect these adventitious sound events. Finally, it is noticeable that the Mel-scaled representation provides the worst results in terms of accuracy, specifically, Acc = 81.5% for wheezes and Acc = 68.7% for crackles. Results suggest that although Mel-scaled spectrogram is well suited to model music and speech signals, the frequencies of interest might not be sufficiently highlighted using this type of spectrogram in the case of adventitious sounds.

In order to test the statistical significance of the results in Fig. 5, we propose to use two of the most widely used and robust non-parametrical tests to compare two sets of distributions, the Man–Whitney U Test and the Wilcoxon signed-rank test [126,127]. Specifically, Table 3 shows the Man–Whitney U Test and Wilcoxon signed-rank test results comparing the Cochleogram, STFT spectrogram and Mel-scaled spectrogram. These tests define both; a null hypothesis (H_o) and an alternative hypothesis (H_1) where the null hypothesis describes the status quo (both results sets compared are statistically equal) and it is assumed to be true unless there are enough arguments to prove the contrary. The p-value determines if we accept or reject the status quo. As can be

[1] https://github.com/loredanadariamang/CODE_EAMBES2022.git.

Table 3
Man–Whitney U Test and Wilcoxon signed-rank test Results for the sets of results obtained in the Fig. 5, using a significance level $\alpha = 0.05$.

Comparison	Mann–Whitney U Test (p-value)	Wilcoxon signed-rank test (p-value)	Significantly better
Crackles			
Cochleogram vs. STFT	2.21e −10	3.78e −09	yes
Cochleogram vs. Mel	1.38e −17	7.55e −10	yes
Wheezes			
Cochleogram vs. STFT	1.91e −06	8.01e −06	yes
Cochleogram vs. Mel	1.49e −17	7.55e −10	yes

Fig. 5. Overall accuracy results evaluating the ICBHI dataset for the Cochleogram (window length of 8 ms), STFT (window length of 32 ms) and Mel-scaled (window length of 32 ms) spectrograms using both the optimal values for the window lengths and overlap sizes. Each box represents 50 data points, each of them associated to a 10-fold cross-validation of the database evaluated. The lower and upper lines of each box show the first and third quartile. The line in the middle of each box represents the median value. The diamond shape in the center of each box represents the average value. The lines extending above and below each box show the extent of the rest of the samples, excluding outliers. Finally, outliers are defined as points that are over 1.5 times the interquartile range from the sample median, which are depicted as crosses.

seen in Table 3, the p-value obtained in all cases does not overpass the significance level $\alpha = 0.05$ used in this analysis, so we can reject the null hypothesis and confirm that the classification performance provided by the Cochleogram is significantly higher compared to both STFT and Mel-scaled spectrograms evaluating both types of adventitious sound such as, crackles and wheezes.

4.2. Four-class normal/crackles/wheezing/both classification results

In this section, we evaluate the performance of the proposed cochleogram representation on a multiclass classification scenario [108] where the aim is to classify a given input breathing sound between healthy (normal respiratory sounds) and unhealthy (crackles, wheezes and crackles+wheezes) respiratory cycle.

Together with the baseline CNN model explained in Section 2.3, we also compare the performance of the different TF representations in Section 2.2 w.r.t a recent approach in [87] which proposes to use artificial noise addition (ANA) by means of an adaptive mechanism of adding a similar type of noise to unhealthy respiratory sounds to enhance the features of ARS and increase the system robustness. Moreover, several state-of-the-art deep learning architectures have been tested including AlexNet [114], ResNet50 [115] and VGG16 [116]. Since the aforementioned deep learning architectures used images as input, we have transformed each computed TF representation matrix into image format by using the Viridis Color Map, which is a homogeneous mapping that utilizes colors changing from blue to green to yellow [80].

Fig. 6 indicates that the use of Cohleagram obtains the best accuracy results using any TF representation and CNN architecture evaluated in this work. In fact, VGG16 provides the best classification performance followed by our baseline CNN model whereas AlexNet and ResNet50 provide similar results in terms of accuracy at the expense of drastically reducing their classification rates. Similar performances were observed in [80] but using pretrained image models from the general purpose Imagenet dataset [128]. On the contrary, in this paper we focus on analyzing the advantages of using alternatives TF representation rather than exploring the possible transfer learning solutions. Regarding the state-of-the-art method in [87], it can be observed that the Spectrogram+ANA strategy obtains competitive results using AlexNet and ResNet50 but its performance drops drastically with respect to the other TF representations when using VGG16 and the Baseline CNN. It suggests that the performance achieved by the Spectrogram+ANA in this work differs from that obtained in [87] probably since the ICBHI database is more complex to analyze because it includes high sound interferences in most respiratory cycles to simulate real acoustic environments.

In Fig. 7, other classification standard metrics are presented including the sensitivity (Fig. 7(a))), specificity (Fig. 7(b))), score (Fig. 7(c))) and precision (Fig. 7(d))). These metrics were introduced in Section 3.1 and provide a further understanding of the methods performance and facilitates the comparison with other approaches evaluated using the ICBHI dataset. Higher values are obtained in terms of specificity (Fig. 7(b)) indicating that all TF representations and CNN architectures evaluated perform better to accurately classify normal (healthy) sounds. Additionally, precision (Fig. 7(d)) and sensitivity values (Fig. 7(a)) report the amount of adventitious sounds correctly classified w.r.t the erroneously classified normal events (precision) and wrongly classified ARS (sensitivity). In this case, the compared methods provide slightly better results in terms of sensitivity which means that, in general, the methods are more reactive to predict adventitious sounds than to classify ARS as normal events. Finally, the score (Fig. 7(c)) represents the average between sensitivity and specificity. In general, it can be observed that the classification behavior of the evaluated TF representations can be considered similar independently of the scenario, binary or four-class, and the deep learning architecture applied as shown in Fig. 5, Fig. 6 and Fig. 7.

In the same way than explained in Section 4.1, a Mann–Whitney U Test and a Wilcoxon signed-rank test were performed, applying a significance level of $\alpha = 0.05$, in order to prove the statistical significance of the Cochleogram as an input of four different CNN architectures in comparison to using spectrograms based on STFT, Mel-scaled and the ANA-based method [87]. As shown in Table 4, results obtained by the Cochleogram indicate a statistic difference with regard to the other studied TF representations for each deep learning architecture so, we can confirm that our proposal based on the Cochleogram significantly improves the learning process of CNN architectures in the field of classification of respiratory sounds compared to the other standard TF representations evaluated.

Fig. 8 shows the accuracy models, loss models and ROC curves for the baseline CNN model using as input data representation the STFT (Figs. 8(a)–8(c)), Mel ((Figs. 8(d)–8(f)), Spectrogram+ANA (Figs. 8(g)–8(i)) and Cochleogram (Figs. 8(j)–8(l)). It can be observed that the behavior of both accuracy and loss models are more stable towards convergence for the Cochleogram in comparison with the rest of the time–frequency representation. In fact, in this work we have used early stopping criteria, to avoid overfitting, accounting to the validation

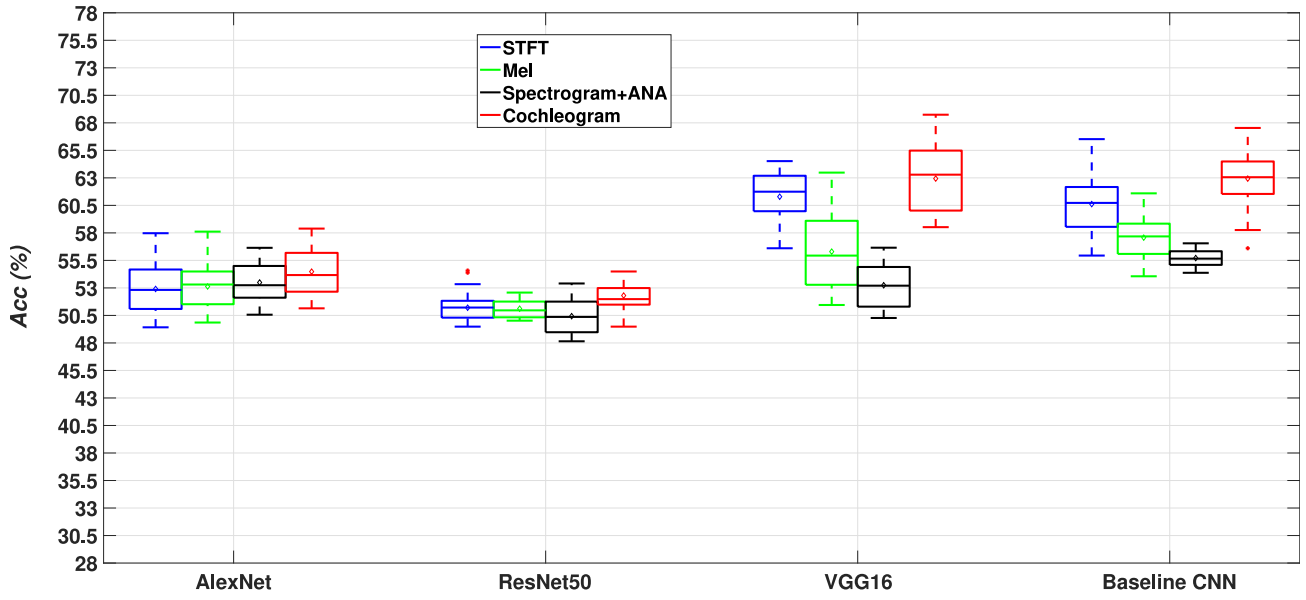

Fig. 6. Performance results, in terms of accuracy, for different CNN networks AlexNet, ResNet50, VGG16 and the CNN implemented in [108] of the STFT spectrogram, Mel-scaled spectrogram, the spectrogram+ANA features in [87] and Cochleogram evaluating four-classes scenario: normal vs. wheezes vs. crackles vs. wheezes+crackles. Each box represents 50 data points, each of them associated to a 10-fold cross validation of the database evaluated. The lower and upper lines of each box show the first and third quartile. The line in the middle of each box represents the median value. The diamond shape in the center of each box represents the average value. The lines extending above and below each box show the extent of the rest of the samples, excluding outliers. Finally, outliers are defined as points that are over 1.5 times the interquartile range from the sample median, which are depicted as crosses.

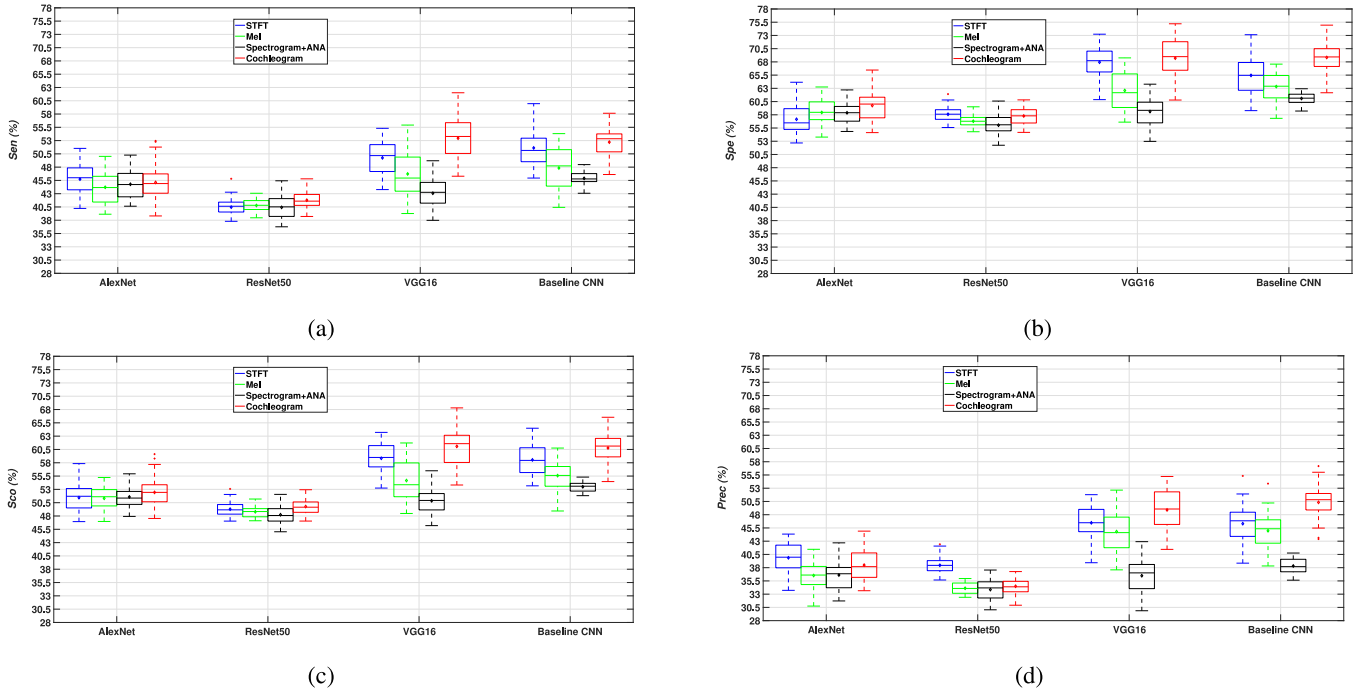

Fig. 7. Performance results of different CNN networks (AlexNet, ResNet50, VGG16 and the CNN implemented in [108]) of the STFT spectrogram, Mel-scaled spectrogram, the Spectrogram+ANA features [87] and Cochleogram evaluating four-classes scenario (normal vs. wheezes vs. crackles vs. wheezes+crackles) in the ICBHI database.

loss parameter which limits the number of epochs in training stage (approximately 16) for the compared methods. As mentioned in [91], this reduced number of epochs could be due to the reduced size of the ICBHI database despite the fact that it is composed of 6898 breathing cycles. Finally, the rightmost subfigure column of Fig. 8 indicates that the Cochleogram can be considered as the most suitable TF representation based on the largest area under the ROC curve which describes quantitatively the classification robustness. This fact, observed in Fig. 8(l), reveals a clearer distinction between all groups of pairs compared to the other time–frequency representations shown in Figs. 8(c), 8(f) and 8(i).

5. Discussion

The ARS classification task has become a challenging research topic in recent years due to growing interest in the promising results obtained using machine learning approaches. However, most of the methods relied on the use of the STFT or the Mel spectrogram to feed data into the

Table 4
Man–Whitney U Test and Wilcoxon signed-rank test Results for the sets of results obtained in the Fig. 6, using a significance level $\alpha = 0.05$.

Comparison	Mann–Whitney U Test	Wilcoxon signed-rank test	Significantly better
AlexNet	(p-value)	(p-value)	
Cochleogram vs. STFT	0.018873	0.002397	yes
Cochleogram vs. Mel	0.004352	0.000374	yes
Cochleogram vs. Spectrogram+ANA	0.020690	0.026728	yes
ResNet50			
Cochleogram vs. STFT	1.94e −05	0.0001595	yes
Cochleogram vs. Mel	4.69e −07	1.74e −06	yes
Cochleogram vs. Spectrogram+ANA	1.13e −07	6.16e −07	yes
VGG16			
Cochleogram vs. STFT	0.006779	0.002555	yes
Cochleogram vs. Mel	1.19e −12	2.66e −09	yes
Cochleogram vs. Spectrogram+ANA	6.94e −18	7.55e −10	yes
Baseline CNN			
Cochleogram vs. STFT	6.94e −05	3.37e −05	yes
Cochleogram vs. Mel	5.39e −14	1.41e −08	yes
Cochleogram vs. Spectrogram+ANA	8.84e −18	7.55e −10	yes

Table 5
Comparison between the proposed method and state-of-the art methods evaluating the four-classes (normal vs. wheezes vs. crackles vs. crackles+wheezes) classification performance in the ICBHI database. Respiratory cycle (RC) represents the temporal length (in seconds) including zero padding to create respiratory cycles of fixed duration. bi-ResNet: bilinear ResNet, NL: non-local, SE: Squeeze-and-Excitation, SA: Spatial Attention, bi-LSTM: bi-directional LSTM, DAG: Directed Acyclic Graph. The rest of the acronyms have been previously mentioned. The references followed by $*$ means that the method has been implemented in this work following the authors description. The results for other methods have been directly extracted from the corresponding works. In bold letter is indicated the maximum value for each metric..

Authors	Time–frequency representation		RC(s)	Technique	Train/Test	Results (%)			
	Type	Parameters				*Sen*	*Spe*	*Sco*	*Acc*
[70]	STFT	30 ms	–	HMM	60/40	–	–	39.6	–
[73]	STFT	500 ms	–	RNN	- (5-fold)	58.4	73.0	65.7	–
[71]	STFT	512 ms	–	HMM SVM	60/40	20.81	78.5	49.65	49.43
[75]	Mel	250 ms	–	RNN	80/20	**64.0**	**84.0**	**74.0**	–
[77]	STFT, Wavelet	20 ms, $D_2 - D_7, A_7$	–	bi-ResNet	- (10-fold)	31.1	69.2	50.2	52.8
[76]	STFT, Scalogram	40 ms	–	CNN	60/40	28.0	81.0	54.0	–
[80]	STFT	64 – 128 – 524 ms	–	CNN SVM	- (10-fold)	–	–	–	65.5
[82]	STFT	20 ms	–	ResNet NL	60/40	41.3	63.2	52.3	–
[79]	Mel	60 ms	–	CNN RNN	80/20	–	58.01	–	–
[83]	STFT	100 ms	2.5	ResNet SE SA	70/30	17.8	81.3	49.6	–
[81]	STFT	–	5	CNN	70/30	–	–	–	74.3
[86]	Mel	–	–	CNN	60/40	–	–	–	**80.4**
[84]	STFT	40 ms	–	CNN bi-LSTM	- (5-fold)	63.0	83.0	73.0	–
[85]	Wavelet	30 ms	–	DAG HMM	–	–	–	–	50.1
[91]	Mel	–	7	CNN	60/40	40.1	72.3	56.2	–
[108]*	STFT	32 ms64 filters	6	CNN	80/20(10-fold)	51.61	65.45	58.53	60.61
	Mel			CNN		47.83	63.33	55.58	57.56
	STFT + Mel			CNN		46.97	63.97	55.47	57.33
[95]	STFT, Log-mel	L32 ms, 50 bins	8	ResNet	60/40	37.2	79.3	58.3	–
[87]*	Spectrogram + ANA	8 ms64 filters	6	CNN (AlexNet)	80/20(10-fold)	44.77	58.37	51.57	53.49
				CNN (ResNet50)		40.42	56.03	48.23	50.43
				CNN (VGG16)		43.08	58.61	50.85	53.24
				CNN (Baseline)		45.88	61.08	53.48	55.71
This work	Cochleagram	84 ms64 filters	6	CNN (AlexNet)	80/20(10-fold)	45.12	59.75	52.43	54.48
				CNN (ResNet50)		41.78	57.78	49.78	52.31
				CNN (VGG16)		53.45	68.71	61.08	62.94
				CNN (Baseline)		52.71	68.84	60.78	62.93

neural network architectures. In this work, we have rigorously studied the effect of using different time–frequency representations and propose the use of the Cochleogram to model the specific temporal and spectral features shown by most of the adventitious respiratory sounds. In addition, a study of the effect of such time–frequency representations on a set of state-of-the-art CNN-based architectures is also presented. Results from the ICBHI database indicate that the Cochleogram, compared to the other evaluated spectrograms, is the most suitable time–frequency representation when it is applied in the task of classifying binary (normal vs. ARS) and four classes of adventitious respiratory sounds (normal vs. crackles vs. wheezes vs. crackles+wheezes).

Table 5 shows a comparison taking into account most of the recent and relevant state-of-the-art methods in the literature for the classification of adventitious respiratory sounds by evaluating the four classes of the ICBHI database. It can be seen that most of these methods use STFT in the preprocessing step in order to compute time–frequency representation of the input data and CNN-based approaches in the classification step.

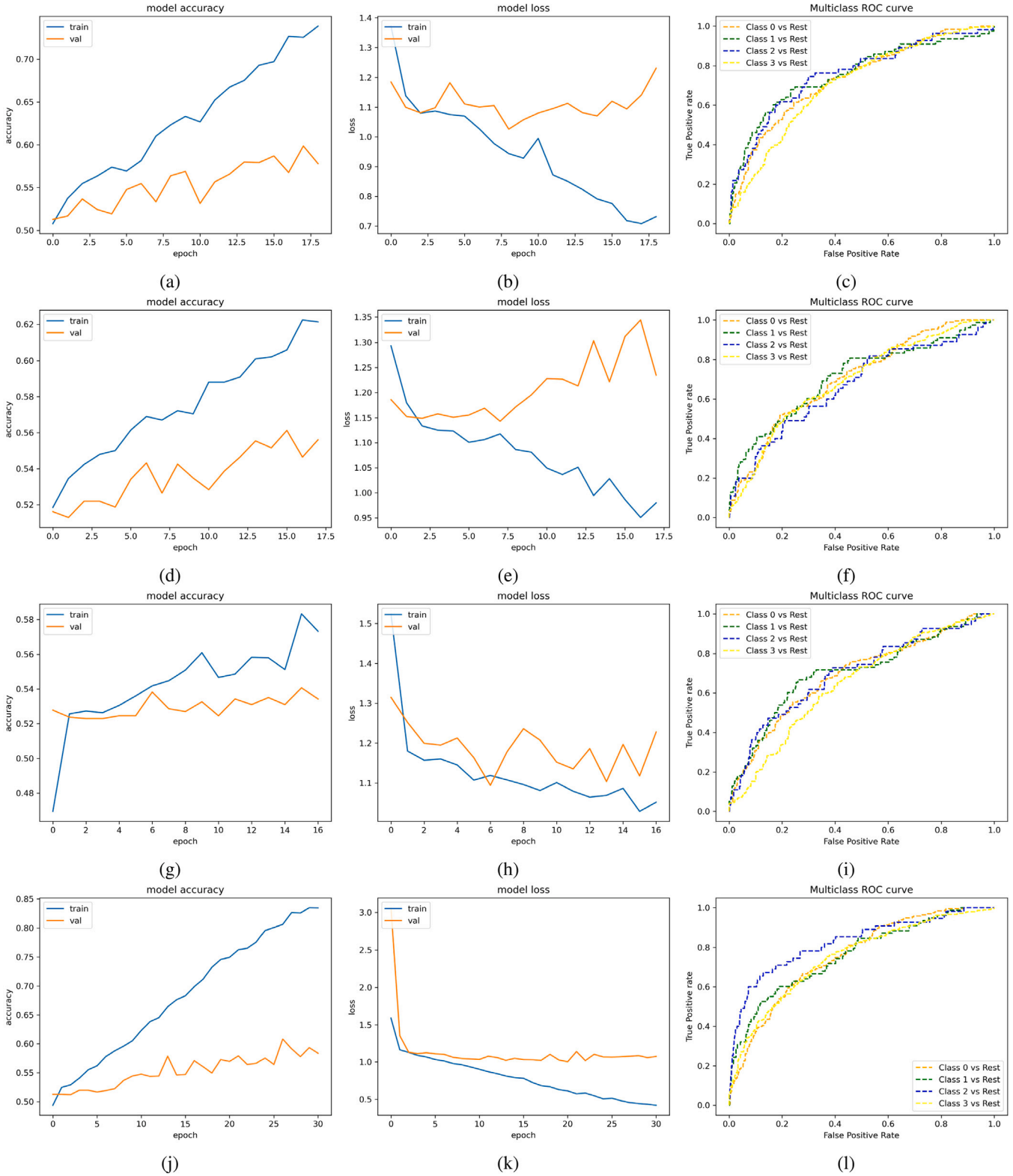

Fig. 8. Model accuracy, model loss and ROC curves when the STFT (a, b, c), Mel (d, e, f), Spectrogram+ANA (g, h, i) and Cochleogram (j, k, l) are used with the Baseline CNN. For each multiclass ROC curve (c, f, i and l), breathing cycles are composed with: Class 0 (crackles), Class 1 (wheezes), Class 2 (crackles+wheezes) and Class 3 (normal).

It should be noted that results show a wide range of performance values due to the lack of uniformity in the evaluation. In fact, some of the methods use only a subset from the ICBHI database [86] instead of using the entire database as we have done in this work. In addition, the metrics used in the evaluation of the different methods differ from one another, which makes it difficult to compare them all [70,79–81,85,86]. With this in mind, the best performance [86] achieves 80.4% in terms of Acc. As for the standard ICBHI database metrics, the best performance is obtained by the Mel+RNN approach [75] achieving Sen = 64.0%, Spe = 84.0% and Sco = 74.0% respectively, which

suggests that there still room for improvement in this biomedical signal processing and machine learning field.

In general terms, it can be observed that Spe results are higher compared to Sen results. It suggests that the generalization of features exhibiting normal respiratory sounds is better learned by the CNN-based model due to the larger number of such sounds existing in the database, which amount to 53% of the entire ICBHI database and 26% more than any other sound so, its feature modeling is more reliable compared to adventitious sounds. In any case, the main purpose of this work is to demonstrate that Cochleagram improves the learning process of a deep learning architecture in the classification of adventitious respiratory sounds.

6. Conclusions and future work

In this paper, we propose the use of the cochleogram-based TF representation to improve the learning process of a CNN model in the classification of respiratory adventitious sounds which has not being applied in this context to the best of our knowledge. Moreover, several TF representations including the STFT spectrogram, Mel-scaled spectrogram the artificial noise addition (ANA) based method in [87] have been considered for comparison. The effect of the different TF representations has been rigorously evaluated using a baseline CNN model for the task of classification of adventitious sounds composed of normal, wheezes, crackles and wheezes+crackles in binary and multiclass scenarios using the largest and most challenging public database of respiratory sounds (ICBHI dataset). Moreover, comparison with several state-of-the-art deep learning models has been performed demonstrating the higher robustness and average accuracy performance of Cochleogram w.r.t the other studied TF representations due to its ability to improve the modeling of respiratory adventitious sounds through more reliable non-uniform spectral resolution and better performance when noise and acoustic changes are active.

Future work will focus on two directions to improve the learning process of deep learning approaches on databases composed of a reduced set of respiratory sounds: (i) novel TF representations to maximize detection/classification performance when simultaneously feeding and integrating dual deep learning architectures; (ii) new approaches combining conventional signal processing techniques that correctly model the spectro-temporal behavior exhibited by specific respiratory sounds, associated with lung pathologies, in cascade with deep learning approaches.

CRediT authorship contribution statement

L.D. Mang: Conceptualization, Methodology, Software, Validation, Formal, Investigation analysis, Writing – original draft, Writing – review & editing. **F.J. Canadas-Quesada:** Conceptualization, Methodology, Software, Validation, Investigation, Resources, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. **J.J. Carabias-Orti:** Conceptualization, Methodology, Software, Validation, Investigation, Resources, Writing – original draft, Writing – review & editing. **E.F. Combarro:** Software, Validation, Investigation analysis, Formal analysis, Project administration, Funding acquisition, Writing – original draft, Writing – review & editing, Supervision. **J. Ranilla:** Software, Validation, Investigation analysis, Formal analysis, Project administration, Funding acquisition, Writing – original draft, Writing – review & editing, Supervision.

Declaration of competing interest

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

Data availability

The database is publicly available (ICBHI: https://bhichallenge.med.auth.gr/ICBHI_2017_Challenge)

References

[1] World Health Organization, The global impact of respiratory diseases (2nd edition), Forum of International Respiratory Societies (FIRS), 2017.

[2] European lung white book, European respiratory society, 2013, http://www.erswhitebook.org/.

[3] Global asthma report. Global asthma network, 2014, http://www.globalasthmareport.org/resources/Global_Asthma_Report_2014.pdf.

[4] C. Cilloniz, I. Martin-Loeches, C. Garcia-Vidal, A. San Jose, A. Torres, Microbial etiology of pneumonia: Epidemiology, diagnosis and resistance patterns, Int. J. Mol. Sci. 17 (12) (2016).

[5] Global Tuberculosis Report 2016, World Health Organization, Geneva, 2016, http://www.who.int/tb/publications/global_report/en/.

[6] L.A. Torre, F. Bray, R.L. Siegel, J. Ferlay, J. Lortet-Tieulent, A. Jemal, Global cancer statistics, 2012, CA: Cancer J. Clin. 65 (2) (2015) 87–108, URL http://www.ncbi.nlm.nih.gov/pubmed/25651787.

[7] J. Torre-Cruz, F. Canadas-Quesada, J. Carabias-Orti, P. Vera-Candeas, N. Ruiz-Reyes, A novel wheezing detection approach based on constrained non-negative matrix factorization, Appl. Acoust. 148 (2019) 276–288.

[8] X.H. Kok, S.A. Imtiaz, E. Rodriguez-Villegas, A novel method for automatic identification of respiratory disease from acoustic recordings, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2019, pp. 2589–2592.

[9] D.E. Bloom, D. Chisholm, E. Jané-Llopis, K. Prettner, A. Stein, A. Feigl, From Burden to "Best Buys": Reducing the Economic Impact of Non-Communicable Disease in Low and Middle-Income Countries, World Health Organization (WHO); World Economic Forum, 2011, URL http://www.who.int/nmh/publications/best_buys_summary.pdf.

[10] A. Sovijarvi, F. Dalmasso, J. Vanderschoot, L. Malmberg, G. Righini, S. Stoneman, Definition of terms for applications of respiratory sounds, Eur. Respir. Rev. 10 (77) (2000) 597–610.

[11] A.J. Salazar, C. Alvarado, F.E. Lozano, System of heart and lung sounds separation for store-and-forward telemedicine applications, in: Revista Facultad de Ingeniería Universidad de Antioquia, (64) Universidad de Antioquia, 2012, pp. 175–181.

[12] J. De La Torre Cruz, F.J. Canadas Quesada, N. Ruiz Reyes, P. Vera Candeas, J.J. Carabias Orti, Wheezing sound separation based on informed inter-segment non-negative matrix partial co-factorization, Sensors 20 (9) (2020) 2679.

[13] A. Sovijarvi, Characteristics of breath sounds and adventitious respiratory sounds, Eur. Respir. Rev. 10 (2000) 591–596.

[14] R.X.A. Pramono, S. Bowyer, E. Rodriguez-Villegas, Automatic adventitious respiratory sound analysis: A systematic review, PLoS One 12 (5) (2017) e0177926.

[15] MedlinePlus. Wheezing, 2016, https://medlineplus.gov/ency/article/003070.htm.

[16] Wheezing causes, Mayo clinic, 2016, https://www.mayoclinic.org/symptoms/wheezing/basics/causes/sym-20050764.

[17] T. Kaisia, A. Sovijärvi, P. Piirilä, H. Rajala, S. Haltsonen, T. Rosqvist, Validated method for automatic detection of lung sound crackles, Med. Biol. Eng. Comput. 29 (5) (1991) 517–521.

[18] K. Zhang, X. Wang, F. Han, H. Zhao, The detection of crackles based on mathematical morphology in spectrogram analysis, Technol. Health Care 23 (s2) (2015) S489–S494.

[19] L. Hadjileontiadis, S. Panas, Nonlinear separation of crackles and squawks from vesicular sounds using third-order statistics, in: Proceedings of 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol. 5, IEEE, 1996, pp. 2217–2219.

[20] S. Charleston-Villalobos, G. Martinez-Hernandez, R. Gonzalez-Camarena, G. Chi-Lem, J.G. Carrillo, T. Aljama-Corrales, Assessment of multichannel lung sounds parameterization for two-class classification in interstitial lung disease patients, Comput. Biol. Med. 41 (7) (2011) 473–482.

[21] L.J. Hadjileontiadis, S.M. Panas, Separation of discontinuous adventitious sounds from vesicular sounds using a wavelet-based filter, IEEE Trans. Biomed. Eng. 44 (12) (1997) 1269–1281.

[22] X. Lu, M. Bahoura, An integrated automated system for crackles extraction and classification, Biomed. Signal Process. Control 3 (3) (2008) 244–254.

[23] G. Serbes, C.O. Sakar, Y.P. Kahya, N. Aydin, Pulmonary crackle detection using time–frequency and time–scale analysis, Digit. Signal Process. 23 (3) (2013) 1012–1021.

[24] P. Stasiakiewicz, A.P. Dobrowolski, T. Targowski, N. Gałązka-Świderek, T. Sadura-Sieklucka, K. Majka, A. Skoczylas, W. Lejkowski, R. Olszewski, Automatic classification of normal and sick patients with crackles using wavelet packet decomposition and support vector machine, Biomed. Signal Process. Control 67 (2021) 102521.

[25] L.J. Hadjileontiadis, Wavelet-based enhancement of lung and bowel sounds using fractal dimension thresholding-Part I: Methodology, IEEE Trans. Biomed. Eng. 52 (6) (2005) 1143–1148.

[26] L.J. Hadjileontiadis, Wavelet-based enhancement of lung and bowel sounds using fractal dimension thresholding-Part II: Application results, IEEE Trans. Biomed. Eng. 52 (6) (2005) 1050–1064.

[27] C. Pinho, A. Oliveira, C. Jácome, J. Rodrigues, A. Marques, Automatic crackle detection algorithm based on fractal dimension and box filtering, Procedia Comput. Sci. 64 (2015) 705–712.

[28] R. Pal, A. Barney, Iterative envelope mean fractal dimension filter for the separation of crackles from normal breath sounds, Biomed. Signal Process. Control 66 (2021) 102454.

[29] X. Liu, W. Ser, J. Zhang, D.Y.T. Goh, Detection of adventitious lung sounds using entropy features and a 2-D threshold setting, in: 2015 10th International Conference on Information, Communications and Signal Processing, ICICS, IEEE, 2015, pp. 1–5.

[30] A. Rizal, R. Hidayat, H.A. Nugroho, Pulmonary crackle feature extraction using tsallis entropy for automatic lung sound classification, in: 2016 1st International Conference on Biomedical Engineering, IBIOMED, IEEE, 2016, pp. 1–4.

[31] L.J. Hadjileontiadis, Empirical mode decomposition and fractal dimension filter, IEEE Eng. Med. Biol. Mag. 26 (1) (2007) 30.

[32] P.A. Mastorocostas, J.B. Theocharis, A dynamic fuzzy neural filter for separation of discontinuous adventitious sounds from vesicular sounds, Comput. Biol. Med. 37 (1) (2007) 60–69.

[33] S.O. Maruf, M.U. Azhar, S.G. Khawaja, M.U. Akram, Crackle separation and classification from normal respiratory sounds using Gaussian mixture model, in: 2015 IEEE 10th International Conference on Industrial and Information Systems, ICIIS, IEEE, 2015, pp. 267–271.

[34] L. Mendes, I.M. Vogiatzis, E. Perantoni, E. Kaimakamis, I. Chouvarda, N. Maglaveras, J. Henriques, P. Carvalho, R.P. Paiva, Detection of crackle events using a multi-feature approach, in: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2016, pp. 3679–3683.

[35] J. Li, Y. Hong, Crackles detection method based on time-frequency features analysis and SVM, in: 2016 IEEE 13th International Conference on Signal Processing, ICSP, IEEE, 2016, pp. 1412–1416.

[36] M. Grønnesby, J.C.A. Solis, E. Holsbø, H. Melbye, L.A. Bongo, Feature extraction for machine learning based crackle detection in lung sounds from a health survey, 2017, arXiv preprint arXiv:1706.00005.

[37] B.A. Pramudita, Istiqomah, A. Rizal, Crackle detection in lung sound using statistical feature of variogram, in: AIP Conference Proceedings, Vol. 2296, (1) AIP Publishing LLC, 2020, 020014.

[38] M. García, S. Villalobos, N.C. Villa, A.J. González, R.G. Camarena, T.A. Corrales, Automated extraction of fine and coarse crackles by independent component analysis, Health Technol. 10 (2) (2020) 459–463.

[39] Y.-X. Liu, Y. Yang, Y.-H. Chen, Lung sound classification based on Hilbert-Huang transform features and multilayer perceptron network, in: 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA ASC, IEEE, 2017, pp. 765–768.

[40] K.J. Hong, S. Essid, W. Ser, D.-G. Foo, A robust audio classification system for detecting pulmonary edema, Biomed. Signal Process. Control 46 (2018) 94–103.

[41] D. Bardou, K. Zhang, S.M. Ahmad, Lung sounds classification using convolutional neural networks, Artif. Intell. Med. 88 (2018) 58–69.

[42] T. Nguyen, F. Pernkopf, Lung sound classification using snapshot ensemble of convolutional neural networks, in: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2020, pp. 760–763.

[43] E. Messner, M. Fediuk, P. Swatek, S. Scheidl, F.-M. Smolle-Juttner, H. Olschewski, F. Pernkopf, Crackle and breathing phase detection in lung sounds with deep bidirectional gated recurrent neural networks, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2018, pp. 356–359.

[44] E. Messner, M. Fediuk, P. Swatek, S. Scheidl, F.-M. Smolle-Jüttner, H. Olschewski, F. Pernkopf, Multi-channel lung sound classification with convolutional recurrent neural networks, Comput. Biol. Med. 122 (2020) 103831.

[45] S.A. Taplidou, L.J. Hadjileontiadis, Wheeze detection based on time-frequency analysis of breath sounds, Comput. Biol. Med. 37 (8) (2007) 1073–1083.

[46] A. Jain, J. Vepa, Lung sound analysis for wheeze episode detection, in: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2008, pp. 2582–2585.

[47] F. Jin, S. Krishnan, F. Sattar, Adventitious sounds identification and extraction using temporal–spectral dominance-based features, IEEE Trans. Biomed. Eng. 58 (11) (2011) 3078–3087.

[48] L. Mendes, I. Vogiatzis, E. Perantoni, E. Kaimakamis, I. Chouvarda, N. Maglaveras, V. Tsara, C. Teixeira, P. Carvalho, J. Henriques, et al., Detection of wheezes using their signature in the spectrogram space and musical features, in: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2015, pp. 5581–5584.

[49] S. Ulukaya, I. Sen, Y.P. Kahya, Feature extraction using time-frequency analysis for monophonic-polyphonic wheeze discrimination, in: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2015, pp. 5412–5415.

[50] R. Naves, B.H. Barbosa, D.D. Ferreira, Classification of lung sounds using higher-order statistics: A divide-and-conquer approach, Comput. Methods Programs Biomed. 129 (2016) 12–20.

[51] S. Cortes, R. Jane, J. Fiz, J. Morera, Monitoring of wheeze duration during spontaneous respiration in asthmatic patients, in: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, IEEE, 2006, pp. 6141–6144.

[52] Y. Qiu, A. Whittaker, M. Lucas, K. Anderson, Automatic wheeze detection based on auditory modelling, Proc. Inst. Mech. Eng. H J. Eng. Med. 219 (3) (2005) 219–227.

[53] S. Le Cam, A. Belghith, C. Collet, F. Salzenstein, Wheezing sounds detection using multivariate generalized Gaussian distributions, in: 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2009, pp. 541–544.

[54] A. Hashemi, H. Arabalibiek, K. Agin, Classification of wheeze sounds using wavelets and neural networks, in: International Conference on Biomedical Engineering and Technology. Vol. 11, (2011) IACSIT Press, Singapore, 2011, pp. 127–131.

[55] S. Ulukaya, G. Serbes, Y.P. Kahya, Wheeze type classification using non-dyadic wavelet transform based optimal energy ratio technique, Comput. Biol. Med. 104 (2019) 175–182.

[56] M. Wisniewski, T.P. Zielinski, Tonality detection methods for wheezes recognition system, in: 2012 19th International Conference on Systems, Signals and Image Processing, IWSSIP, IEEE, 2012, pp. 472–475.

[57] M. Wiśniewski, T.P. Zieliński, Joint application of audio spectral envelope and tonality index in an e-asthma monitoring system, IEEE J. Biomed. Health Inf. 19 (3) (2014) 1009–1018.

[58] M. Bahoura, C. Pelletier, Respiratory sounds classification using Gaussian mixture models, in: Canadian Conference on Electrical and Computer Engineering. Vol. 3, IEEE, 2004, pp. 1309–1312.

[59] P. Mayorga, C. Druzgalski, R. Morelos, O. Gonzalez, J. Vidales, Acoustics based assessment of respiratory diseases using GMM classification, in: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, IEEE, 2010, pp. 6312–6316.

[60] J. Zhang, W. Ser, J. Yu, T. Zhang, A novel wheeze detection method for wearable monitoring systems, in: 2009 International Symposium on Intelligent Ubiquitous Computing and Education, IEEE, 2009, pp. 331–334.

[61] M. Bahoura, Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes, Comput. Biol. Med. 39 (9) (2009) 824–843.

[62] B.-S. Lin, H.-D. Wu, S.-J. Chen, Automatic wheezing detection based on signal processing of spectrogram and back-propagation neural network, J. Healthc. Eng. 6 (4) (2015) 649–672.

[63] K. Kochetov, E. Putin, S. Azizov, I. Skorobogatov, A. Filchenkov, Wheeze detection using convolutional neural networks, in: EPIA Conference on Artificial Intelligence, Springer, 2017, pp. 162–173.

[64] D. Oletic, V. Bilas, Asthmatic wheeze detection from compressively sensed respiratory sound spectra, IEEE J. Biomed. Health Inf. 22 (5) (2017) 1406–1414.

[65] J. Torre-Cruz, F. Canadas-Quesada, S. García-Galán, N. Ruiz-Reyes, P. Vera-Candeas, J. Carabias-Orti, A constrained tonal semi-supervised non-negative matrix factorization to classify presence/absence of wheezing in respiratory sounds, Appl. Acoust. 161 (2020) 107188.

[66] J.D.L.T. Cruz, F.J.C. Quesada, J.J.C. Orti, P.V. Candeas, N.R. Reyes, Combining a recursive approach via non-negative matrix factorization and Gini index sparsity to improve reliable detection of wheezing sounds, Expert Syst. Appl. 147 (2020) 113212.

[67] B.M. Rocha, D. Filos, L. Mendes, G. Serbes, S. Ulukaya, Y.P. Kahya, N. Jakovljevic, T.L. Turukalo, I.M. Vogiatzis, E. Perantoni, et al., An open access database for the evaluation of respiratory sound classification algorithms, Physiol. Meas. 40 (3) (2019) 035001.

[68] ICBHI 2017 challenge, respiratory sound database, 2017, https://bhichallenge.med.auth.gr/ICBHI_2017_Challenge.

[69] G. Serbes, S. Ulukaya, Y.P. Kahya, An automated lung sound preprocessing and classification system based on spectral analysis methods, in: International Conference on Biomedical and Health Informatics, Springer, 2017, pp. 45–49.

[70] N. Jakovljević, T. Lončar-Turukalo, Hidden markov model based respiratory sound classification, in: International Conference on Biomedical and Health Informatics, Springer, 2017, pp. 39–43.

[71] G. Chambres, P. Hanna, M. Desainte-Catherine, Automatic detection of patient with respiratory diseases using lung sound analysis, in: 2018 International Conference on Content-Based Multimedia Indexing, CBMI, IEEE, 2018, pp. 1–6.

[72] M. Aykanat, Ö. Kılıç, B. Kurt, S. Saryal, Classification of lung sounds using convolutional neural networks, EURASIP J. Image Video Process. 2017 (1) (2017) 1–9.

[73] K. Kochetov, E. Putin, M. Balashov, A. Filchenkov, A. Shalyto, Noise masking recurrent neural network for respiratory sound classification, in: International Conference on Artificial Neural Networks, Springer, 2018, pp. 208–217.

[74] R. Liu, S. Cai, K. Zhang, N. Hu, Detection of adventitious respiratory sounds based on convolutional neural network, in: 2019 International Conference on Intelligent Informatics and Biomedical Sciences, ICIIBMS, IEEE, 2019, pp. 298–303.

[75] D. Perna, A. Tagarelli, Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks, in: 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems, CBMS, IEEE, 2019, pp. 50–55.

[76] K. Minami, H. Lu, H. Kim, S. Mabu, Y. Hirano, S. Kido, Automatic classification of large-scale respiratory sound dataset based on convolutional neural network, in: 2019 19th International Conference on Control, Automation and Systems, ICCAS, IEEE, 2019, pp. 804–807.

[77] Y. Ma, X. Xu, Q. Yu, Y. Zhang, Y. Li, J. Zhao, G. Wang, LungBRN: A smart digital stethoscope for detecting respiratory disease using bi-resnet deep learning algorithm, in: 2019 IEEE Biomedical Circuits and Systems Conference, BioCAS, IEEE, 2019, pp. 1–4.

[78] D. Ngo, L. Pham, A. Nguyen, B. Phan, K. Tran, T. Nguyen, Deep learning framework applied for predicting anomaly of respiratory sounds, in: 2021 International Symposium on Electrical and Electronics Engineering, ISEE, IEEE, 2021, pp. 42–47.

[79] J. Acharya, A. Basu, Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning, IEEE Trans. Biomed. Circuits Syst. 14 (3) (2020) 535–544.

[80] F. Demir, A.M. Ismael, A. Sengur, Classification of lung sounds with CNN model using parallel pooling structure, IEEE Access 8 (2020) 105376–105383.

[81] A. Saraiva, D. Santos, A. Francisco, J. Sousa, N. Ferreira, S. Soares, A. Valente, Classification of respiratory sounds with convolutional neural network, in: Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies - BIOINFORMATICS, SciTePress. INSTICC, 2020, pp. 138–144.

[82] Y. Ma, X. Xu, Y. Li, LungRN+ NL: An improved adventitious lung sound classification using non-local block ResNet neural network with mixup data augmentation, in: Interspeech, 2020, pp. 2902–2906.

[83] Z. Yang, S. Liu, M. Song, E. Parada-Cabaleiro, B.W. Schuller, Adventitious respiratory classification using attentive residual neural networks, 2020.

[84] N. Asatani, T. Kamiya, S. Mabu, S. Kido, Classification of respiratory sounds using improved convolutional recurrent neural network, Comput. Electr. Eng. 94 (2021) 107367.

[85] S. Ntalampiras, I. Potamitis, Automatic acoustic identification of respiratory diseases, Evol. Syst. 12 (1) (2021) 69–77.

[86] H. Chanane, M. Bahoura, Convolutional neural network-based model for lung sounds classification, in: 2021 IEEE International Midwest Symposium on Circuits and Systems, MWSCAS, IEEE, 2021, pp. 555–558.

[87] R. Zulfiqar, F. Majeed, R. Irfan, H.T. Rauf, E. Benkhelifa, A.N. Belkacem, Abnormal respiratory sounds classification using deep CNN through artificial noise addition, Front. Med. 8 (2021).

[88] B. Abdelkader, S. Ouhbi, A. Lakas, E. Benkhelifa, C. Chen, End-to-end AI-based point-of-care diagnosis system for classifying respiratory illnesses and early detection of COVID-19, Front. Med. (2021).

[89] Y. Kim, Y. Hyon, S.S. Jung, S. Lee, G. Yoo, C. Chung, T. Ha, Respiratory sound classification for crackles, wheezes, and Rhonchi in the clinical field using deep learning, Sci. Rep. 11 (1) (2021) 1–11.

[90] W. Song, J. Han, H. Song, Contrastive embeddind learning method for respiratory sound classification, in: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2021, pp. 1275–1279.

[91] S. Gairola, F. Tom, N. Kwatra, M. Jain, Respirenet: A deep neural network for accurately detecting abnormal lung sounds in limited data setting, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2021, pp. 527–530.

[92] A. Srivastava, S. Jain, R. Miranda, S. Patil, S. Pandya, K. Kotecha, Deep learning based respiratory sound analysis for detection of chronic obstructive pulmonary disease, PeerJ Comput. Sci. 7 (2021) e369.

[93] Z. Tariq, S.K. Shah, Y. Lee, Feature-based fusion using CNN for lung and heart sound classification, Sensors 22 (4) (2022) 1521.

[94] Y. Choi, H. Choi, H. Lee, S. Lee, H. Lee, Lightweight skip connections with efficient feature stacking for respiratory sound classification, IEEE Access (2022).

[95] T. Nguyen, F. Pernkopf, Lung sound classification using co-tuning and stochastic normalization, IEEE Trans. Biomed. Eng. (2022).

[96] Z. Zhao, Z. Gong, M. Niu, J. Ma, H. Wang, Z. Zhang, Y. Li, Automatic respiratory sound classification via multi-branch temporal convolutional network, in: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2022, pp. 9102–9106.

[97] J. Saldanha, S. Chakraborty, S. Patil, K. Kotecha, S. Kumar, A. Nayyar, Data augmentation using Variational Autoencoders for improvement of respiratory disease classification, PLoS One 17 (8) (2022) e0266467.

[98] G. Petmezas, G.-A. Cheimariotis, L. Stefanopoulos, B. Rocha, R.P. Paiva, A.K. Katsaggelos, N. Maglaveras, Automated lung sound classification using a hybrid CNN-LSTM network and focal loss function, Sensors 22 (3) (2022) 1232.

[99] N. Gajhede, O. Beck, H. Purwins, Convolutional neural networks with batch normalization for classifying hi-hat, snare, and bass percussion sound samples, in: Proceedings of the Audio Mostly 2016, 2016, pp. 111–115.

[100] S. Kwon, A CNN-assisted enhanced audio signal processing for speech emotion recognition, Sensors 20 (1) (2019) 183.

[101] J. Kim, A.-D. Nguyen, S. Lee, Deep CNN-based blind image quality predictor, IEEE Trans. Neural Netw. Learn. Syst. 30 (1) (2018) 11–24.

[102] C. Tian, Y. Xu, Z. Li, W. Zuo, L. Fei, H. Liu, Attention-guided CNN for image denoising, Neural Netw. 124 (2020) 117–129.

[103] U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, M. Adam, A. Gertych, R. San Tan, A deep convolutional neural network model to classify heartbeats, Comput. Biol. Med. 89 (2017) 389–396.

[104] N. Baghel, M.K. Dutta, R. Burget, Automatic diagnosis of multiple cardiac diseases from PCG signals using convolutional neural network, Comput. Methods Programs Biomed. 197 (2020) 105750.

[105] S.B. Shuvo, S.N. Ali, S.I. Swapnil, T. Hasan, M.I.H. Bhuiyan, A lightweight cnn model for detecting respiratory diseases from lung auscultation sounds using emd-cwt-based hybrid scalogram, IEEE J. Biomed. Health Inf. 25 (7) (2020) 2595–2603.

[106] S. Jayalakshmy, G.F. Sudha, Scalogram based prediction model for respiratory disorders using optimized convolutional neural networks, Artif. Intell. Med. 103 (2020) 101809.

[107] M. Fraiwan, L. Fraiwan, M. Alkhodari, O. Hassanin, Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory, J. Ambient Intell. Humaniz. Comput. (2021) 1–13.

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

[109] C. Roads, The Computer Music Tutorial, MIT Press, 1996.

[110] S. Schulz, T. Herfet, Binaural source separation in non-ideal reverberant environments, in: Proceedings of 10th International Conference on Digital Audio Effects (DAFx-07), Bordeaux, France, 2007.

[111] J. Chen, Y. Wang, D. Wang, A feature study for classification-based speech separation at low signal-to-noise ratios, IEEE/ACM Trans. Audio Speech Lang. Process. 22 (12) (2014) 1993–2002.

[112] R.V. Sharan, T.J. Moir, Acoustic event recognition using cochleagram image and convolutional neural networks, Appl. Acoust. 148 (2019) 62–66.

[113] S. Das, S. Pal, M. Mitra, Acoustic feature based unsupervised approach of heart sound event detection, Comput. Biol. Med. 126 (2020) 103990.

[114] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25 (2012).

[115] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, 2015.

[116] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.

[117] S. Davis, P. Mermelstein, Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences, IEEE Trans. Acoust. Speech Signal Process. 28 (4) (1980) 357–366.

[118] D. O'shaughnessy, Speech Communications: Human and Machine, IEEE, Addison- Wesley Pub. Co. 1987.

[119] N. Nakamura, M. Yamashita, S. Matsunaga, Detection of patients considering observation frequency of continuous and discontinuous adventitious sounds in lung sounds, in: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2016, pp. 3457–3460.

[120] L. Xu, J. Cheng, J. Liu, H. Kuang, F. Wu, J. Wang, ARSC-net: Adventitious respiratory sound classification network using parallel paths with channel-spatial attention, in: 2021 IEEE International Conference on Bioinformatics and Biomedicine, BIBM, IEEE, 2021, pp. 1125–1130.

[121] H. Purwins, B. Li, T. Virtanen, J. Schlüter, S.-Y. Chang, T. Sainath, Deep learning for audio signal processing, IEEE J. Sel. Top. Sign. Proces. 13 (2) (2019) 206–219.

[122] R.D. Patterson, I. Nimmo-Smith, J. Holdsworth, P. Rice, An efficient auditory filterbank based on the gammatone function, in: A Meeting of the IOC Speech Group on Auditory Modelling At RSRE. Vol. 2. No. 7, 1987.

[123] X. Valero, F. Alias, Gammatone cepstral coefficients: Biologically inspired features for non-speech audio classification, IEEE Trans. Multimed. 14 (6) (2012) 1684–1689.

[124] B. Gao, W.L. Woo, L. Khor, Cochleagram-based audio pattern separation using two-dimensional non-negative matrix factorization with automatic sparsity adaptation, J. Acoust. Soc. Am. 135 (3) (2014) 1171–1185.

[125] D. Berrar, Cross-validation, in: S. Ranganathan, M. Gribskov, K. Nakai, C. Schönbach (Eds.), Encyclopedia of Bioinformatics and Computational Biology, Academic Press, Oxford, 2019, pp. 542–545, URL https://www.sciencedirect.com/science/article/pii/B978012809633820349X.

[126] F. Wilcoxon, Individual comparisons by ranking methods, in: Breakthroughs in Statistics, Springer, 1992, pp. 196–202.

[127] H.B. Mann, D.R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, Ann. Math. Stat. (1947) 50–60.

[128] 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, pp. 248–255.