G Model
ARTMED-1591; No. of Pages 12
ARTICLE IN PRESS
Artificial Intelligence in Medicine xxx (2018) xxx–xxx
Contents lists available at ScienceDirect
Artificial Intelligence in Medicine
journal homepage: www.elsevier.com/locate/aiim

Lung sounds classification using convolutional neural networks

Dalal Bardou [a,*], Kun Zhang [a,*], Sayed Mohammad Ahmad [b]

[a] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China
[b] Lareb Technologies, India

ARTICLE INFO

Article history:
Received 5 May 2017
Received in revised form 18 April 2018
Accepted 23 April 2018

Keywords:
Convolutional neural network
Lung sounds classification
Handcrafted features extraction
Deep learning
Models ensembling
Support vector machines

ABSTRACT

In recent years, the classification of lung sounds has been the topic of interest in the field of bioinformatics. Lung sounds convey relevant information related to pulmonary disorders, and to evaluate patients with pulmonary conditions, the physician or the doctor uses the traditional auscultation technique. However, this technique suffers from limitations. For example, if the physician is not well trained, this may lead to a wrong diagnosis. Moreover, lung sounds are non-stationary, complicating the tasks of analysis, recognition, and distinction. This is why developing automatic recognition systems can help to deal with these limitations. In this paper, we compare three machine learning approaches for lung sounds classification. The first two approaches are based on the extraction of a set of handcrafted features trained by three different classifiers (support vector machines, k-nearest neighbor, and Gaussian mixture models) while the third approach is based on the design of convolutional neural networks (CNN). In the first approach, we extracted the 12 MFCC coefficients from the audio files then calculated six MFCCs statistics. We also experimented normalization using zero mean and unity variance to enhance accuracy. In the second approach, the local binary pattern (LBP) features are extracted from the visual representation of the audio files (spectrograms). The features are normalized using whitening. The dataset used in this work consists of seven classes (normal, coarse crackle, fine crackle, monophonic wheeze, polyphonic wheeze, squawk, and stridor). We have also experimentally tested dataset augmentation techniques on the spectrograms to enhance the ultimate accuracy of the CNN. The results show that CNN outperformed the handcrafted feature based classifiers.

1. Introduction

Lung sound characteristics and their diagnoses form an indispensable part of pulmonary pathology [1,2]. Auscultation is a technique whereby physicians evaluate and diagnose patients with pulmonary conditions by using a stethoscope. It is known to be inexpensive, non-invasive, and safe, besides taking less time for diagnosis [1,2]. It also provides much information about the respiratory organ and the signs of the diseases that affect it [3]. However, if not done by a well-trained physician, this may lead to wrong diagnosis.

Lung sounds being non-stationary signals, it is both difficult to analyze and hard to distinguish them with traditional auscultation methods. Thus, the use of an electronic stethoscope, coupled with a pattern recognition system, helps to overcome the limitations of traditional auscultation; and provides an efficient method for clinical diagnosis [4,5].

Lung sounds can be divided into two categories: normal breathing sounds or adventitious breathing sounds. Normal breathing sounds are heard when no respiratory disorders exist and adventitious sounds are heard when a respiratory disorder does exist [6,7]. Normal respiratory sounds consist of tracheal, bronchial and bronchovesicular sounds. On the chest wall, a normal respiratory sound is characterized by a low noise during inspiration, and a hardly audible noise during expiration. On the trachea, a normal respiratory sound is characterized by a broader spectrum of noise, such as a noise containing higher-frequency components, which is audible during both the inspiratory and expiratory phase [8]. An adventitious sound is an additional respiratory sound that is superimposed onto normal breath sounds. These can be continuous, like wheezes, or discontinuous, like crackles. Some of them, like squawks, have both characteristics. The presence of such sounds usually indicates a pulmonary disorder [8].

Crackles are discontinuous and explosive adventitious sounds. They appear much more during the inspiratory phase. They are characterized by their specific waveform, their duration, and their location in the respiratory cycle. A crackle can be characterized by its total duration: fine crackles have a short duration and coarse crackles have a long duration [8]. Fine crackles are present in high

* Corresponding authors.
E-mail addresses: dalal.bardou@njust.edu.cn (D. Bardou), zhangkun@njust.edu.cn (K. Zhang).

https://doi.org/10.1016/j.artmed.2018.04.008
0933-3657/

frequencies. Crackle occurs in many diseases, such as congestive heart failure, pneumonia, bronchiectasis, pulmonary fibrosis, and chronic diffuse parenchymal lung disease. The American Thoracic Society [9] have defined fine crackles as having an initial deflection width (IDW) of 0.7 ms and 2 cycle durations (2CD) of 5 ms, and coarse crackles as having an IDW of 1.5 ms and 2CD of 10 ms. Another group has defined fine crackles as having a 2CD < 10 ms, and coarse crackles as having a 2CD > 10 ms [10]. Additionally, the frequency spectrum of crackles is between 200 Hz and 2000 Hz [11,12].

Wheezes are high-pitched continuous adventitious sounds. A wheeze can be monophonic if it contains a single frequency, or polyphonic if several frequencies are simultaneously perceived [11]. Rhonchi are low-pitched, continuous sounds, and they are characterized by a dominant frequency of about 200 Hz or less [5]. The diseases associated with wheezing sounds are asthma, pneumonia, and bronchitis.

There are many other categories of lung sounds, including pleural rub, stridor, and squawks. A pleural rub is the characteristic sound produced when inflamed pleural surfaces rub together during respiration; a stridor is a very loud wheeze; a squawk is a short inspiratory wheeze [13].

In this paper, we have proposed three classification approaches to classifying seven types of lung sounds. The lung sounds consist of normal sounds, coarse crackle, fine crackle, monophonic wheeze, polyphonic wheeze, stridor, and squawk. The objectives of our study are: **(a)** to compare the handcrafted features based classification methods and the convolutional neural networks, **(b)** to test the power and the suitability of the convolutional neural networks to address lung sounds classification task, **(c)** to put to the test dataset normalization and augmentations influence on the performances, and **(d)**, to compare between handcrafted features extracted directly from audio files and the ones extracted from the visual representation of audio files (spectrograms).

The rest of the paper is organized as follows: in Section 2, we give an overview of the previous studies related to lung sounds classification using handcrafted features based classification methods, and the recent convolutional neural networks applications related to medical sounds classification as well. In Section 3, we give information about the original data and explained how the data set is constructed and preprocessed, while Section 3 is dedicated to the data augmentation techniques applied to spectrograms. In Section 4, we give the topology of the proposed CNN. In Section 5, we describe the handcrafted features-based classification, the classifiers, as well as the implementations settings. Finally, in Section 6 and 7, performances and experimental results are provided, and a comparison with the literature is made.

2. Related work

With the onset of pattern recognition and artificial intelligence, many feature-based approaches have been proposed, to develop automatic systems for the classification of different lungs sounds. The support vector machine (SVM) is known to be a promising method, and so it has been used to address this task. In [4], the authors used the frequency ratio of power spectral density (PSD) values, and the Hilbert-Huang transform (HHT) features, to distinguish between normal lung sounds, crackles, and rhonchus. The SVM classifier achieved an accuracy of above 90%. In [14], time-frequency (TF) and time-scale (TS) analysis are proposed for the detection of pulmonary crackles. In the feature extraction step, the frequency characteristics of crackles, using TF and TS analysis, are extracted from both the non-pre-processed and pre-processed signals. For pre-processing, DTCWT is applied, with the aim of removing the frequency bands that do not contain crackle information. In classification step, k-Nearest Neighbors (k-NN), SVM, and multilayer perceptron are used to classify crackling and non-crackling sounds. The SVM classifier achieved the best result, obtaining a classification accuracy of 97.5%. In [15], SVM is used to classify respiratory sounds, ranging from normal to continuous adventitious, including wheezing, stridor, and rhonchi. The removed features consisted of instantaneous kurtosis, discriminating function, and entropy. The optimally achieved classification accuracy was between 97.7% and 98.8%. In [16], another approach is proposed for the classification of normal and continuous adventitious signals, although only wheeze signals were used as continuous adventitious signals. The mel-frequency cepstral coefficients (MFCCs) are used for feature extraction, the gaussian mixture model (GMM) is used to classify the signals, and the achieved accuracy was 94.2%. In [17], a technique to obtain the time-frequency representation (TFR) of thoracic sounds is proposed. Using TF patterns, the authors assessed the performance of the TFRs for the heart, adventitious, and normal lung sounds. After simulations, they concluded that the best TFR performance was achieved by the Hilbert–Huang spectrum (HHS). In [5], the classification of lung sounds using higher order statistics (HOS) is proposed. In feature extraction step, HOS is used to extract features (second, third, and fourth order cumulants) from five types of lung sounds (normal, coarse crackle, fine crackle, monophonic and polyphonic wheezes). Genetic algorithms and Fisher's discriminant ratio are used to reduce dimensionality, and for classification, k-NN, and naive Bayes classifiers are used to classify lung sound events in a tree-based system. The classifier accuracy was 98.1% accuracy on training data, and 94.6% on validation data. In [18], the authors used MFCC features along with artificial neural network (ANN) to classify normal sounds, wheezes, and crackles. The classification achievement performance was 75% for crackles, 100% for wheezes and 80% for normal. The authors, in comparison with previous studies in [19,20], concluded that GMM is 15% more accurate than ANN for crackle classification. In [21], the authors proposed a method for the separation and classification of crackles from normal respiratory sounds, using GMM. This work consists of four steps: preprocessing, feature extraction, feature selection and classification. In the preprocessing step, a band-pass filter is used for background noises reduction, and then three spatial-temporal features, namely pitch, energy, and spectrogram, were extracted. After the feature selection step, the final features are trained using GMM. The achieved accuracy was 97.56%. In [22], the authors proposed a novel attractor recurrent neural networks (ARNN) topology, based on the fuzzy functions (FFs-ARNN), for the classification of lung abnormalities. The respiratory sounds are modelled using an ARNN and a FFs-ARNN, and to evaluate their performances, recurrent quantification analysis (RQA) was used. The best accuracy was 91%, which was achieved using FFs-ARNN with RQA features. While in [23], the authors propose a new recurrent fuzzy filter, based on a pipelined Takagi–Sugeno–Kang for a real-time separation of discontinuous adventitious sounds (DAS) and vesicular sounds (VS), in [24], an orthogonal least squares-based fuzzy filter (OLS-FF) is proposed for the same task. Two fuzzy inference systems are used in parallel to perform the task of adaptive separation, resulting in the OLS-FF. In [25], the authors proposed an automatic method for the elimination of vesicular sound from crackle signal, with minimal distortion of crackle parameters. After selecting a region of interest, distortion metric, based on the correlation between raw and filtered waveforms in that region, is defined. In [26], the authors proposed signal processing methodologies for the detection of crackles in audio files. After the extraction of a window of interest, based on fractal dimension and box filtering, the potential crackle is verified and validated, and crackle parameters are extracted and characterized.

The convolutional neural networks (CNN) can be regarded as a variant of the standard neural networks. Instead of using fully con-