Gabriela Augustinov, Patrick Fischer, Volker Gross, Ulrich Koehler, Keywan Sohrabi, and Seyed Amir Hossein Tabatabaei*

Automatic Detection and Classification of Cough Events Based on Deep Learning

Abstract: In this paper, a deep learning approach for classification of cough sound segments is presented. The architecture of the network is based on a pre-trained network and the spectrogram images of three recording channels have been extracted for the sake of training the network. The classification accuracy based on three recording channels is 92% for a binary classification model and the network converges fast. Two classification models based on binary and multi-class problems are proposed. Relevant classification parameters including the Receiver Operating Characteristic (ROC) curve are reported.

Keywords: Deep Learning, Convolutional Neural Networks, Respiratory sounds, Classification, Spectrogram

https://doi.org/10.1515/cdbme-2020-3083

1 Introduction and Related Work

1.1 Introduction

Pervasive analysis of respiratory acoustic sounds including adventitious sounds via artificial intelligence has gained enormous attention during the past years. The analysis consists of automatic detection and classification of adventitious respiratory lung sounds via machine learning methods. There exist few taxonomies for such sound types. In fact, the modern framework for the classification of respiratory lung sound has been developed by the Computerized Respiratory Sound Analysis (CORSA) group in the early 2000s. This framework defines all sound phenomena categorized as breath sounds or adventitious respiratory sounds. Adventitious respiratory sounds are introducing additional noise detected by auscultation and are further subdivided into continuous and discontinuous adventitious respiratory sounds. The continuous subset is mainly indicated by wheezing, snoring and stridor. Prominent examples of discontinuous subset are crackle and cough. Our paper addresses the problem of automatic detection of cough events in respiratory sound recordings. The problem of automatic detection and possible classification of respiratory sounds is a highly challenging problem. The main challenges are lack of enough relevant acoustic data, special difficulties in data acquisition, ethical, privacy and security issues and data cleansing which targets noise and redundancy removal. The presented work in this paper focuses mainly on the automatic detection of cough events in Chronic Obstructive Pulmonary Disease (COPD) patients based on a machine learning approach. The proposed machine leaning approach uses a deep learning structure for the detection of healthy respiratory recordings from cough holder segments. The main advantage of using a deep learning structure over other methods is that it waives the need for handcrafting feature engineering on the acquired data. The proposed approach in this paper uses the spectrograms of the recorded acoustic sounds as the inputs to the deep structure. The features will be learned through the depth of network based on the fed data. The rest of this paper is as follows. The required background and related work are presented in the following subsection. Section II describes the structure of the deep learning network. The description of the experiments and the discussion are presented in Section III. Section IV concludes the paper.

1.2 Background and Related Work

A classical approach towards adventitious respiratory event detection and classification based on machine learning starts

***Corresponding author:** Seyed Amir Hossein Tabatabaei, e-mail: amir.taba1.61@gmail.com

Gabriela Augustinov, Volker Gross, Keywan Sohrabi: Faculty of Health Sciences, University of Applied Sciences, Giessen, Germany

Patrick Fischer, Seyed Amir Hossein Tabatabaei: Institute of Medical Informatics, Justus-Liebig University Giessen, Germany

Ulrich Koehler: Department of Internal Medicine, Pneumology, Intensive Care and Sleep Medicine, University Hospital of Marburg and Giessen, Marburg, Germany

like other approaches with data recording (data acquisition) and pre-processing. The pre-processing of such acoustic data is of special sensitivity as they are contaminated with noises and redundant private and public pieces like speech and environmental voice. There are some methods in the literature like speech obfuscation techniques addressing the latter issue [1, 2]. Following the classical approach, the feature engineering step is performed on the pre-processed and cleansed acoustic data. Mel Frequency Cepstral Coefficients (MFCCs) [3, 4] and Linear Predictive Coding (LPC) [5] are traditional acoustic features commonly used in acoustic data analysis for many years. Using image features in acoustic data detection and classification is another approach in the feature engineering process. For example, data represented by the image spectrograms of the recorded acoustic signals in data acquisition can be pre-processed prior to feature extraction [6]. A comprehensive list of the used features in acoustic analysis of pulmonary diagnostic together with their performance and use-cases is given in [7]. A range of simple or ensemble classifiers from Support Vector Machine (SVM), Artificial Neural Networks (ANNs), Hidden Markov Models (HMM), random forest, decision trees to gradient boosting models have been engaged for binary or multi-class classification tasks. However, to avoid challenges and difficulties concerning feature engineering in the classification pipeline, deep learning methods have been utilized. The deep learning structures in acoustic data analysis let statistical model and data-extracted filters to be learned directly based on the raw waveform of the recorded signals [4]. Raw waveform extracted from multiple recording channels can be the input to Recurrent Neural Networks (RNNs) wherein the temporal dependency between the sequences are taken into account. A survey of the relative works on the acoustic methods for pulmonary diagnosis has been presented in [7]. For example, as a relevant work, the proposed method in [6] introduces two classification models based on deep learning wherein two CNN and RNN-based deep structures are utilized. The spectrogram of the recording segment is used as the input data to the networks. The CNN-based network employs a sequence of two convolutional-max-pooling layers followed by a fully-connected layer to perform a three-label classification task. The classification accuracy of 87.6% and 79.7% have been reported for CNN-based and RNN-based networks respectively. In another work presented in [8], a deep neural network is engaged in order to detect cough holder segments from healthy segments. The reported average accuracy is 92.3% while the average sensitivity and specificity are 97.6% and 93.7% respectively. The proposed work in [9] uses Principal Component Analysis (PCA) and a deep learning network to detect and classify the cough sounds as productive and non-productive as well as the ambient sounds based on the signal spectra. So, a three-label classification problem is addressed in which the accuracy of 99.91% is achieved.

2 Structure of the Proposed Deep Learning Network

2.1 Network Architecture

As the aim of this work is to evaluate the detecting capability of a deep learning architecture with respect to visual data, the acoustic recorded data are presented by two-dimensional spectrogram images. The spectrogram images are extracted by applying Fourier transform on windowed, zero-padded signals. Windows length has been chosen to be a power of 2 in order to apply Fast Fourier Transform (FFT) algorithm. We also introduced a 50 % window overlap. Afterwards the default color scale has been adjusted for better visual interpretability. The deep architecture here engages the CNN as its building blocks. So, the input size as well as the resolution of the spectrograms remain unchanged to adapt to the network adjusted parameters. The respiratory sound segments with the short recording time will be padded by zeros just in case to adapt to the fixed input size of the network. The classification in our approach is a twofold classification problem. At first a binary classification wherein a sequence to sequence labelling will be performed is presented. The second classification type is a 3-label classification in which segments with no-cough, one-cough and multiple-coughs are classified. For the two-class problem the spectrograms are labelled as cough or non-cough in which a cough segment might contain one or multiple cough events. The trained classifier assigns a label to the spectrograms from the test/evaluation set. We have used a CNN-based pre-trained network for the sake of transfer learning. The selected network is AlexNet [10] which is suitable for multi-label classification as well. The architecture of our utilized network includes five convolutional layers consisting of ReLU as the activation function and Max pooling operation. The layers are followed by three fully connected layers consisting of the same activation function with dropout. The final Softmax function is adjusted according to the type of classification problem. The input image data to the network is an RGB image of size 227 x 227 pixels per channel taken by the first convolutional

layer where zero-center normalization and zero padding also take place. The filter in the first layer is of size 11 x 11 with stride size of 4 x 4 followed by a Maxpooling of size 3 x 3. Each image corresponds to a recording microphone channel and three channels as left, right and trachea have been used for this sake. The fully connected layer performs regularization of rate 0.5 to avoid Gradient vanishing or overfitting phenomenon. The detailed architecture of the network is summarized in Table 1. In this table, Conv(n_1, n_2, n_3) states for an n_1 filtering operation of size n_2 by n_2 and stride size of n_3 by n_3. Also Max(n_1, n_2) states for a Maxpooling operation of size n_1 and stride size of n_2 by n_2.

Table 1. Structure of the utilized network

ConvLayer 1	Conv(96, 11, 4)	Max(3,2)
ConvLayer 2	Conv(2×128, 5, 1)	Max(3,2)
ConvLayer 3	Conv(384, 3, 1)	----------
ConvLayer 4	Conv(2×192, 3, 1)	----------
ConvLayer 5	Conv(2×128, 3, 1)	Max(3,2)
FCLayer 6	Dropout = 0.5	----------
FCLayer 7	Dropout = 0.5	Softmax

2.2 Data Acquisition

Data acquisition has been performed by using the LEOSound Lung-Sound-Monitor (Loewenstein Medical GmbH & Co. KG, Bad Ems, Germany). LEOSound is a class I certified medical device for recording and detecting several lung sounds. Data is captured by three bioacoustics sensors, which are placed on the neck (paralaryngeal) and back of the patient (see Fig. 1 in [11]). All recorded data are stored in binary format with a sampling rate of 5512 Hz and 16-bit depth. The recorded data is split into 30-sec windows, which is a common approach in sleep medicine and which has been transferred to LEOSound as well. The training data recordings have been recorded in two consecutive nights in patients suffering from Chronic Obstructive Pulmonary Disease (COPD). In total 48 patients were measured for at least 8 hours, resulting in 100,507 windows of 30-second length. Data labelling was performed in a similar fashion for each dataset, where medical experts scanned through the recordings to identify cough events and also rated whether the respective event was productive or not.

2.3 Network Training

The collected data for training and testing the network consist of 1602 spectrogram images where 801 samples are cough holder segments and the rest are healthy, resulting to a fully balanced dataset. The train set includes 80% of the total set and the test set consists of the rest of 20%. The usage of CNN structure results in an elaborate feature extraction in the classification model. The training process for a 3-label classification problem is performed similarly to the 2-label classification by using a balanced data set consisting of a total of 945 images in which one-third of the dataset is dedicated to each class label. The labels are healthy, one-cough and multiple-cough. The examples of high-level features extracted from the utilized network are shown in Figure 1.

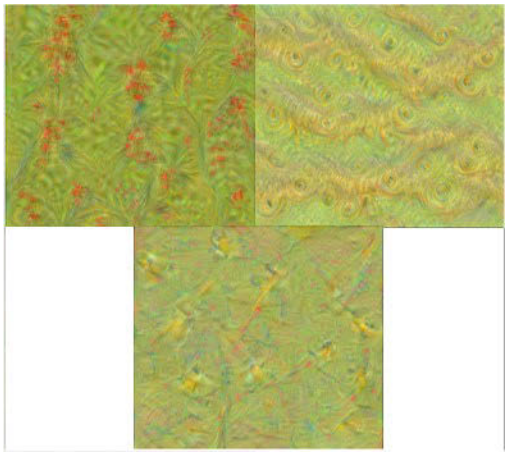

Figure 1. The high-level features of a multiple-cough (top left), one-cough (top right) and no-cough class (bottom)

The utilized objective function in the network training is Stochastic Gradient Descent with Momentum. The initial learning rate is set to 0.001. The employed network is trained and converges within 20 epochs in the two-label case. The models have been implemented via MATLAB deep learning toolbox on a single NVIDIA GPU with 8 GB RAM resource.

3 Results and Discussion

Common classification parameters including accuracy, sensitivity, specificity and Receiver Operating Characteristic (ROC) curve have been extracted for the classification performance in both cases. The aforementioned parameters for both classification problems are shown in Table 2 below.

Table 2. Classification parameters of the models

Classification type	Sensitivity	Specificity	Accuracy
2-label	98%	87%	92%
3-label	72%, 81%, 89%	89%, 86%, 96%	81%

The order of measures in the three-label classification model is corresponding to one-cough, multiple-cough and no-cough labels respectively. The precision and F1-measure score as a harmonic mean of the corresponding precision and recall (sensitivity) indicating their trade-off are calculated for the 2-label classification as 88% and 92% respectively. Similar measures for the 3-label classification are 77%, 74%, 92% and 74%, 77%, 90% respectively. The Area Under the Curve (AUC) for the ROC curves for both classification models is 0.97 and 0.92 respectively. It is worth to mention that the area under the ROC curve for the second model is estimated by averaging the three corresponding areas under the ROCs. Also, the confusion matrices of both classification problems are shown as follows.

Table 3. Confusion matrix of both classification models

Observed				
Output			Cough	No-cough
		Cough	157	21
		No-cough	3	139
Output		One-cough	Multiple-cough	No-cough
	One-cough	91	19	8
	Multiple-cough	30	103	5
	No-cough	5	4	113

4 Conclusion

In this paper, a machine learning approach for cough sound detection has been presented. The proposed solution is based on a deep learning architecture wherein the elaborate feature selection process is performed via CNN automatically. The reported accuracy is 92.5% in the binary classification model however it decreases in the 3-label classification model to 81.2%. In future, we will utilize deep learning through CNN/LSTM fed by raw signal data with short time steps. Also, we will carry out event localization in which the fine time interval of the event will be localized. The classification of different cough types and cough alongside other sounds will be another future track. Finally, incorporating data of patients with comorbidities would be planned as well.

Author Statement

Research funding: The authors state no funding involved. Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent is not applicable.

Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee.

References

[1] S. Lee, E. Nemati and J. Kuang, „Configurable Pulmonary-Tuned Privacy Preservation Algorithm for Mobile Devices," in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Madrid, Spain, 2018.

[2] X. Sun, Z. Lu, W. Hu and G. Cao, „SymDetector: Detecting sound-related respiratory symptoms using smartphones," in *UbiComp 2015 - Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2015.

[3] S. Furui, „Speaker-independent isolated word recognition based on emphasized spectral dynamics," in *ICASSP*, 1986.

[4] H. Purwins, B. Li, T. Virtanen, J. Schlüter, S. Chang and T. Sainath, „Deep Learning for Audio Signal Processing," *Journal of Selected Topics of Signal Processing,* Bd. 13, Nr. 2, pp. 206-219, 2019.

[5] L. Deng and D. O'Shaughnessy, „Speech Processing: A Dynamic and Optimization-Oriented Approach" , New York: Marcel Dekker, INC, pp. 41-48, 2003.

[6] J. Amoh and K. Odame, „Deep neural networks for identifying cough sounds," *IEEE Transactions on Biomedical Circuits and Systems,* Bd. 10, Nr. 5, pp. 1003-1011, 2016.

[7] A. Rao, E. Huynh, TJ. Royston, A. Kornblith and S. Roy, „Acoustic methods for pulmonary diagnosis," *IEEE Reviews in Biomedical Engineering,* Bd. 12, pp. 221-239, 2019.

[8] P. Kadambi, A. Mohanty, H. Ren, J. Smith, K. McGuinnes, K. Holt, A. Furtwaengler, R. Slepetys, Z. Yang, J. Seo, J. Chae, Y. Cao and V. Berisha, „Towards a Wearable Cough Detector Based on Neural Networks," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018.

[9] S. Khomsay, R. Vanijjirattikhan and J. Suwatthikul, „Cough detection using PCA and Deep Learning," in *2019 International Conference on Information and Communication Technology Convergence (ICTC)*, 2019.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, „ImageNet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.,* pp. 1-9, 2012.

[11] B. M. Rocha, D. Filos, L. Mendes, I. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Jácome, A. Marques, R. P. Paival, . Chouvarda, P. Carvalho and N. Maglaveras, „A respiratory sound database for the development of automated classification," *Precision Medicine Powered by pHealth and Connected Health,* pp. 33-37, 2018.