Wheeze Detection Using Convolutional Neural Networks

Kirill Kochetov[1(✉)], Evgeny Putin[1], Svyatoslav Azizov[1], Ilya Skorobogatov[2], and Andrey Filchenkov[1]

七次ACM世界冠军

[1] Computer Technologies Lab, ITMO University, 49 Kronverksky Pr, 197101 St. Petersburg, Russia
{kskochetov,eoputin,szazizov,afilchenkov}@corp.ifmo.ru

[2] Center for Billing Technologies and Printing Services, 17 Bolshaya Raznochinaya St, 197110 St. Petersburg, Russia

Abstract. In this paper, we propose to use convolutional neural networks for automatic wheeze detection in lung sounds. We present convolutional neural network based approach that has several advantages compared to the previous approaches described in the literature. Our method surpasses the standard machine learning models on this task. It is robust to lung sound shifting and requires minimal feature preprocessing steps. Our approach achieves 99% accuracy and 0.96 AUC on our datasets.

Keywords: Wheeze detection · Convolutional neural networks · Machine learning · Deep learning

1 Introduction

According to the World Health Organization (WHO), lung diseases are the third most common cause of the death right after the coronary heart disease and stroke. Prevalence statistics of the pulmonary diseases are impressive. For example, in many countries, approximately 5% of the population suffers from asthma, which is appearing as coughing, dyspnea, and the main factor is wheezes.

Wheezes are adventitious sounds present in the lung that is clinically defined as abnormal. Wheezes can be determined as an undesigned and uninterrupted sounds [1]. Lung sound records generated during breathing can be a good source of information for lung treatment. For example, the presence of wheezes in lung sound records of children have been widely used as a parameter to evaluate the inclination to asthma. From an acoustic point of view, wheezes are characterized by periodic waveforms with a dominant frequency usually over 100 Hz [2].

Traditionally, a stethoscope is used to diagnose and monitor wheezes. Stethoscope is a fast, reliable, non-invasive instrument for diagnosing respiration functions of patients. But due to the fact of the increasing number of patients with asthma and other lung diseases, there is a permanent demand for automatic wheeze detection systems. Among other reasons that complicate lung treatment,

E. Oliveira et al. (Eds.): EPIA 2017, LNAI 10423, pp. 162–173, 2017.
DOI: 10.1007/978-3-319-65340-2_14

maybe, the most important one is the late diagnosis. In many countries with low quality of life, there are simply not enough qualified medical workers to diagnose every patient on time. Thus, for patients with lung diseases uninterrupted and automatic monitoring is very important as the day-to-day state monitoring of lung health can provide key information to the medical diagnosis. The automated wheeze recognition system will also allow any member of medical staff to understand if something is wrong with the respiratory cycle of patient, so it would speed up further treatment. Such systems could help to minimize or eliminate human factor mistakes serving as intellectual decision support systems for medical workers.

The scientific advances in signal processing, speech recognition, time series analysis has been very significant lately, so there are a lot of complex and powerful methods to analyze respiratory cycle sounds nowadays. However, the necessary quality level has not been achieved yet preventing widespread of such systems in real clinics. One of the main problems of most previous studies in wheeze detection is that almost all the methods used there require a lot of complex preprocessing steps of lung sound records to achieve suitable performance. These preprocessing steps are not robust and are sensitive to internal/external noise and quality of records. For example, conventional machine learning (ML) models such as support vector machine (SVM) or k-nearest neighbors (kNN), if being trained on mel-frequency cepstral coefficients (MFCC) features, will work poorly on shifted lung sounds, because MFCC features by nature are not robust to shifting. But in real clinical applications, it is hard to adjust records so they start and end exactly when they are required to. Also, ML models trained on preprocessed data in such ways may not generalize their performance on different versions of stethoscopes.

Deep learning models [3] have recently showed very promising performance on a range of tasks. It has turned out to be a very efficient tool for image recognition [4–6], nature language processing [7–9] and speech recognition [10–13]. In 2012, a convolutional neural network (CNN) was trained to classify 1.3 million high-resolution images into the 1000 different classes [14]. It achieved top-1 and top-5 error rates of 39.7% and 18.9% on the test data. In the area of sound event recognition [15], CNNs beats other feature extraction algorithms and showed great performance in terms of robustness compared with MFCC.

The goal of this study is suggesting an efficient approach for the wheeze detection problem. We propose approach wheeze detection using convolutional neural network (WDCNN) that solves this problem using more flexible architecture. The proposed approach reaches state-of-the-art performance in wheeze detection task and is robust for shifting records and external noise. The contribution of this paper includes:

这是最高纪录了！！！！

第一篇用cnn的；
raw data preprocessing
性能最好；最鲁棒

crackles

- To the best of our knowledge, we are the first who presents CNN-based automatic wheeze detection approach.
- Our CNN-based approach on non-processing normalized data like spectrograms are better in terms of evaluation metrics than other approaches that involved a lot of complex preprocessing steps like applying different filters

(FIR, Hamming window, etc.), noise reduction techniques, normalization by frequencies, different feature extraction techniques (MFCC, SBC, Entropy features), etc.
– Our CNN-based solution achieves 99% accuracy and 0.96 AUC measure.

The rest of the paper is organized as follows. In Sect. 2, the previous studies related to the topic are summarized. We describe our deep learning approach, data that we used and methods for preprocessing in Sect. 3. Experiments details are presented in Sect. 4. Results and performance measures are presented in Sect. 5. Finally, conclusions are made in Sect. 6.

2 Related Work

In [16], authors proposed an approach to automatic wheeze detection called Entropy-Based Wheeze Detection (EBWD). On the first stage, digital signal was transformed to domain frequency by commonly used STFT (Short Time Fourier Transform) method. STFT procedure produced the Fourier spectrum for each shorter segment of digital signal. On the second stage, peak detection by masking was applied to identify peaks in the signal. After that the authors empirically estimated that vacancy areas beside the peaks for wheezy breath were much larger than for normal breath. They characterized these discrepancies in signals in terms of informational entropy. On the third stage, several features were extracted with respect to entropy. The authors proposed to use features like the difference between maximum and minimum entropy or their ratio and to perform thus wheeze detection using some threshold on this features. As a result, EBWD was able to identify 85% of wheeze samples based on Microphones (Panasonic WM-64 ON) data.

The authors of another paper [17] have demonstrated how Gaussian Mixture Models (GMM) can be applied to classify respiratory diseases. RALE database and 4 classes (normal, wheeze, crackles and asthma) of lung sounds were used in this research. As the preprocessing steps for digital signals from a stethoscope, the authors applied FIR (Finite Impulse Response) filter followed by FFT (Fast Fourier Transform) algorithm. After that, MFCC (Mel-frequency cepstral coefficients) features were extracted from preprocessed signals. The authors conducted several experiments to evaluate the best number of GMMs in the model and also evaluated the accuracy of their models with and without cross-validation. As a result, their approach obtained average (across all classes) 98.7% accuracy without cross-validation and 52.5% accuracy with cross-validation.

In [18], the authors have used a lot of machine learning methods such as feed-forward multilayer neural network (MLP), random forest (RF), logistic regression (LR), naïve Bayes (NB), support vector machine (SVM) and k-nearest neighbors (kNN) for automated wheeze detection using phonopneumograms from the Internet (called INT dataset) and Dubrovnik General Hospital (DGH) datasets. Their preprocessing stages consisted of applying Yule-Walker filter to reduce the influence of cardiovascular and muscular noise followed by STFT procedure. After that, the commonly used MFCC features were extracted from the

preprocessed signals. Also, the authors experimented with some statistical features using FFT (Renyi entropy, Kurtosis, Spectral Flatness (SF), Skewness, etc.) without using MFCC features. As a result, their approach on statistical features achieved 93.62%, 91.77% accuracies for INT and DGH datasets, respectively, by the best model (NN with 2 hidden layers). On MFFC features, their approach with SVM and kNN models obtained 99% on both datasets. The authors claimed that by properly filtering and preprocessing the entry data, and using MFCC features, the signals recorded in suboptimal conditions can be efficiently processed.

Table 1. The comparison table

Author	Year	Processing and feature extraction methods	Classifier	Result
Zhang et al.	2009	Entropy, STFT	Entropy-Based Wheeze Detection (EBWD)	Able to identify 85% wheezes samples and systems have been implemented into wearable sound-based respiration monitoring system
Mayorga et al.	2010	Finite Impuse Response (FIR) filter, Hamming window, Fast Fourier Transform (FFT), Mel Frequency Cepstral Coefficients (MFCC)	Gaussian Mixture Models (GMM)	Accuracy of 98.7% obtained
Milicevic et al.	2016	Yule-Walker filter, STFT, FFT, MFCC	MLP, RF, LR, NB, SVM, kNN	93.62%, 91.77% accuracies for INT and DGH datasets, respectively
Bahoura et al.	2004	MFCC, Subband Based Cepstral (SBC)	Gaussian Mixture Models (GMM), MLP, Vector Quantization (VQ)	GMM on MFCC successfully classify sounds into two category (wheeze and normal sounds)
Palaniappan et al.	2014	MFCC	SVM, kNN	92.19% and 98.26% accuracies obtained for SVM and kNN, respectively

GMM method was used for binary classification of normal and wheezing respiratory sounds in [19] as well as in [18]. At the preprocessing stage, the sound data were divided into overlapped frames from which a reduced dimension feature vectors were extracted using MFCC procedure and Subband based Cepstral parameters (SBC). The GMM-MFCC combination was compared with Vector

Quantization (VQ) and Multi-Layer Perceptron (MLP) neural networks and the best result was obtained by GMM-MFCC. At the postprocessing stage, smoothing of the score function was applied to include the wheezes duration into the model. This led to the significant performance improvement.

In [20], two common classifiers were applied to the data of three categories: normal, airway obstruction pathology, and parenchymal pathology. The MFCC features were extracted from the data and analyzed by one-way ANOVA (analysis of variance). After that, the features were fed separately into the SVM and kNN classifiers. The obtained accuracies of the classification with the SVM and kNN classifiers were found to be 92.19% and 98.26%, respectively.

The most successful recent papers related to wheeze detection are listed in Table 1. These works presented effective techniques that achieved high accuracy values of the exploited classification methods. The Table reflects differences between the techniques, specifically, it shows which processing, feature extraction methods, and classifiers were applied in each work and to what results it has led.

3 Proposed Approach

3.1 Data

Most authors use lung sound data from different internet databases (e.g. [21]) for their research purposes. Such databases were created foremost for educational aims. Each database usually contains a small amount of samples. Unfortunately, most of these databases consist almost only of sick lung sounds.

We collected several datasets from all over the Internet (e.g. soundtracks from YouTube videos and a other sources, e.g. [21]). Some of founded lung sound samples were too long or too short. The number of such samples was 43. These samples were considered as outliers and were dropped. Resulting dataset consisted of 817 samples with 232 healthy and 585 sick lung respiratory cycles. A 2D spectrograms of several samples are shown in Fig. 1.

3.2 Methods

There are many preprocessing techniques and feature extraction methods that can be used for lung sound data. Some of these methods were used as a part of the methodology in the previous studies, which were described in Sect. 2. Most of the observed methodologies included many complex preprocessing steps [22] followed by feature extraction methods like MFCC. There is a good explanation of how MFCC works in [23] A lot of the proposed preprocessing techniques were explained by the fact that there were many data sources with different distributions of frequencies, noises, contents and other characteristics of lung sound data. A methodology has to be insensitive to data changing, data shifting or to some anomalies in the lung sounds. But most of the previously proposed methods were just data cleansing and normalization techniques required to apply

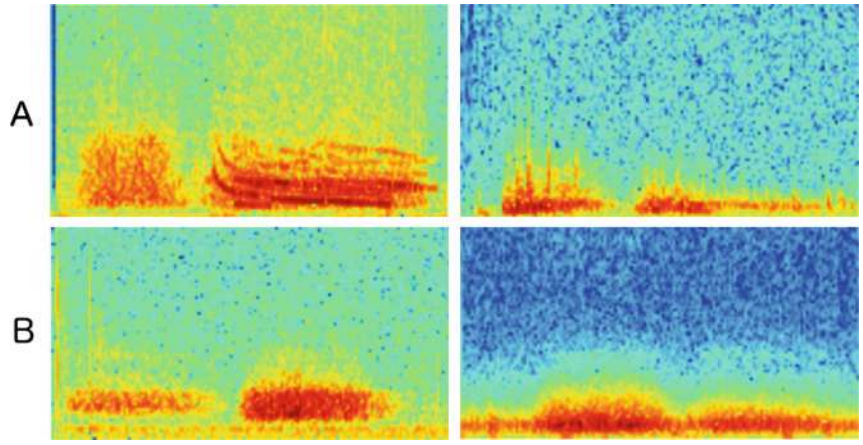

Fig. 1. Examples of respiratory cycle spectrograms. (A) With anomalies. (B) Without anomalies

machine learning models. For example, finite impulse response (FIR) filter [17] and Yule-Walker filter [18] were used for reducing the impact of noise and for the same frequency range in generated spectrograms by STFT. A classifier required clean normalized data to achieve the better performance score on validation data and to give more stable predictions on the future lung sounds (preprocessed in the same way). In this section, we describe some techniques used in our deep learning approach or used in approaches that we implemented for comparison purposes.

Our deep learning approach for wheeze detection is short and clear. We use a minimum of preprocessing steps and give to CNN almost raw lung sound data. Study design of our approach and CNN scheme is presented in Fig. 2.

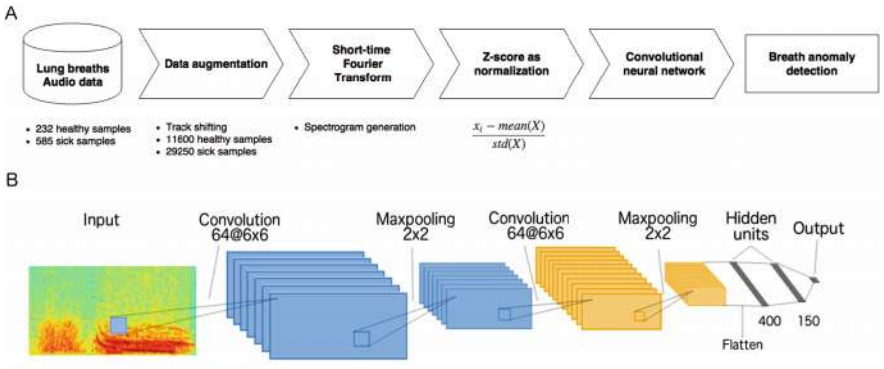

Fig. 2. Deep learning approach. (A) Study design. (B) CNN architecture

网上有augmentor！

Data Augmentation is a technique for artificially enlargement of the original data. Deep learning models (such as CNN) require as much data as possible, thus data augmentation is very helpful here to generate new samples and it is also applied to prevent overfitting. Also, due to data augmentation, we can show CNN robustness in terms of respiratory cycle shifting (it is presented in Sects. 4 and 5 of this paper). Here we consider shifting operation as the biasing of the original lung sound sample by several frames in time.

这是什么技术？

Data augmentation is the first step of our approach and it is applied to the original data. The soundtracks containing respiratory cycles are shifted several times and are marked by id of the original soundtrack. Marking is required to avoid intersection of train and test sets during model training.

音轨变换技术=data augmentor

The soundtrack shifting technique helps to simulate real-world conditions and to generate many possible variations of specific respiratory cycle. If a classifier can precisely detect anomalies regardless of location of content in the lung sound sample, it is robust enough. For example, a patient can make a mistake and turn on the electronic stethoscope in the middle of the inspiration or turn it off before the end of the cycle.

STFT. Short-time Fourier transform (STFT) is a general-purpose tool for audio signal processing. It is a time-dependent Fourier transform for a sequence, and it is computed using a sliding window. The STFT is a Fourier-related transform that is used to determine the sinusoidal frequency and the phase content of the local sections of a signal as it changes over time.

$$\mathrm{STFT}\{x[n]\}(m,\omega) \equiv X(m,\omega) = \sum_{n=-\infty}^{\infty} x[n] \times w[n-m] \times e^{-j\omega n} \tag{1}$$

where $w[n]$ is the window (frame) and $x[n]$ is the signal to be transformed. The magnitude squared of the STFT yields the spectrogram of the function:

$$\mathrm{spectrogram}(t,\omega) = |\mathrm{STFT}(t,\omega)|^2 \tag{2}$$

which is represented like plots as shown in Fig. 1. We use STFT in our approach for obtaining a raw spectrogram of lung sound.

***Z*-Score.** It is a normalization technique using formula presented on Fig. 3. This type of normalization applies scale on each feature (pixel) of the sample. If we do not scale the input vectors, the ranges of the features (pixels) would likely be different, and this may cause problems related to the specific of gradient based algorithms.

把特征向量中心化

CNN. Convolutional neural network is a type of feed-forward artificial neural network most commonly used for image processing. There is a good explanation of how CNN and it layers works in [24]. Briefly, CNN consist of convolution, pooling and fully connected layers. Convolution layer is a feature extraction part of CNN. It learns filters that are activated when they detect some specific

feature (shape) on some position in the input. Max pooling layer is a form of down-sampling. It converts the input image into a set of non-overlapping rectangles and, for each such sub-region, outputs the maximum. Fully connected layers followed by convolutional and pooling layers processes high-level abstract features received by learnable convolutions. Also fully connected layer is a main building block of regular feed-forward neural network, in which all neurons have full connections to all activations in the previous layer. The output of CNN depends on the task. In our wheeze detection task it is binary classification problem and there is one neuron in the output of CNN.

4 Experiments

4.1 Experiments Design

In this study, we conducted three types of experiments, the only difference in which is the data augmentation application.

A. A common experiment was conducted on non-augmented data for comparison. Original dataset without data augmentation was used for this experiment. The purpose of this experiment is to compare CNN and other ML models in terms of classification performance. 1.无data augmentation的实验做增强；
B. Each soundtrack was augmented 50 times and then the data were splitted into the train and test sets. There are no augmented copies of any soundtracks from the train set in the test set. The purpose of this experiment is to test CNN and other models for robustness in terms of soundtrack shifting. 2.测试集的增强数据和训练集的增强数据没overlap，用来测robustness
C. This experiment is similar to "B". The only difference is that there are no augmented samples in the train set. The purpose of this experiment is to test whenever or not our CNN-based model holds its generalization ability on shifted samples trained on the original ones. 3.训练集里没有增强的数据！

All experiments were conducted on a computer with Intel Core i7-6900 CPU with 32 GB of RAM and NVIDIA 1080 GPU.

4.2 Result Evaluation

Due to the unbalanced dataset, we used Area Under Curve (AUC) and Matthews correlation coefficient (MCC) as the performance measures.

AUC score is an area under ROC (receiver operating characteristic) curve. ROC is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. The curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings.

MCC is a robust measure for the binary classification task in the case of very different sizes of classes. It returns a value between -1 and $+1$. A coefficient of $+1$ represents an ideal prediction, 0 corresponds to no better than a random prediction and -1 indicates the total disagreement between prediction and

不平衡数据集可以用马太相关系数（类似于打分）和AUC；及交叉验证

observation. MCC is calculated using Eq. 3. In this equation, TP is the number of true positives, TN the number of true negatives, FP the number of false positives and FN the number of false negatives.

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{3}$$

Cross-validation was used to evaluate the results. The idea behind cross-validation is to divide the data set into disjoint training and validation subsets K in different but regular ways, after that a performance measure is evaluated as the mean value on all folds. Thus, results from cross-validation experiments are robust. We used 5-fold cross-validation.

5 Results

5.1 Results of the Experiments

The classification performance of different models trained on MFCC features is presented in Table 2. The best result on non-augmented data (MCC 0.88) was achieved by CNN, but MLP, RF and GBM scored just a bit worse (MCC 0.824). In the second (B) and third (C) experiments, CNN completely beat other ML methods in terms of performance and robustness. CNN completely outperformed other models on augmented data, especially when tested on augmented data (0.897 versus 0.32 MCC). CNN robustness in this task is explained by the specificity of model structure. Convolutional layers consist of many learnable filters (feature masks). Filters are independent of content (respiratory cycle) location. Because of this, filters can detect anomalies regardless of soundtrack shifting.

增强以后cnn会打败其他方法；

可学习的滤波器，滤波器与内容位置无关；

Additionally, Fig. 3 shows the performance of classifiers in the form of ROC curves for experiment “B”. By analyzing the shown curves, as well as the corresponding AUC score, it can be seen that the best results are achieved by CNN.

Also, we provide results of CNN trained on almost raw features (spectrograms). The results are presented in Table 3. As expected, CNN showed a good feature extraction ability. Our model showed better performance in experiment “B” because of ability of deep learning models to process and memorize big amount of data. Also, we provide loss curve of the best trained CNN for experiment “B”. It is shown on Fig. 3.

Comparison with State-of-the-Art Models. Almost all the approaches overviewed in Sect. 2 contained MFCC as the feature extraction step. Performance of our approach is 96% of accuracy on MFCC features and 99% on spectrograms. The best accuracy scores were achieved in [17] (98.7%) and in [20] (98.26%). But almost all of the experiments conducted in observed papers used a small data set and were not tested for algorithm robustness.

基于频谱比基于mfcc好！

Table 2. Classification performance received with MFCC features. (A) Original train and test. (B_1) Augmented train and test. (B_2) Augmented train and original test. (C) Original train and augmented test

	A		B_1		B_2		C	
Model	AUC	MCC	AUC	MCC	AUC	MCC	AUC	MCC
SVM	0.787	0.523	0.557	0.095	0.731	0.418	0.5	0
KNN	0.631	0.356	0.531	0.158	0.625	0.331	0.509	0.026
LR	0.79	0.546	0.599	0.134	0.784	0.515	0.505	0.01
GBM	0.874	0.824	0.808	0.279	0.823	0.769	0.507	0.053
RF	0.874	0.824	0.744	0.207	0.77	0.729	0.506	0.091
MLP	0.874	0.824	0.729	0.32	0.866	0.819	0.507	0.061
CNN	**0.939**	**0.88**	**0.939**	**0.897**	**0.931**	**0.83**	**0.723**	**0.519**

支持向量机
k-聚类
逻辑回归

Table 3. Classification performances of CNN on spectrogram and MFCC features. (MFCC) Augmented train and test. MFCC features were used. (A) Original train and test. Here and below spectrogram features were used. (B_1) Augmented train and test. (B_2) Augmented train and original test. (C) Original train and augmented test

Experiment	Accuracy	F1	AUC	MCC
MFCC	0.961	0.956	0.939	0.897
A	0.989	0.981	0.958	0.916
B_1	0.984	0.977	0.95	0.915
B_2	**0.99**	**0.982**	**0.96**	**0.922**
C	0.904	0.856	0.772	0.546

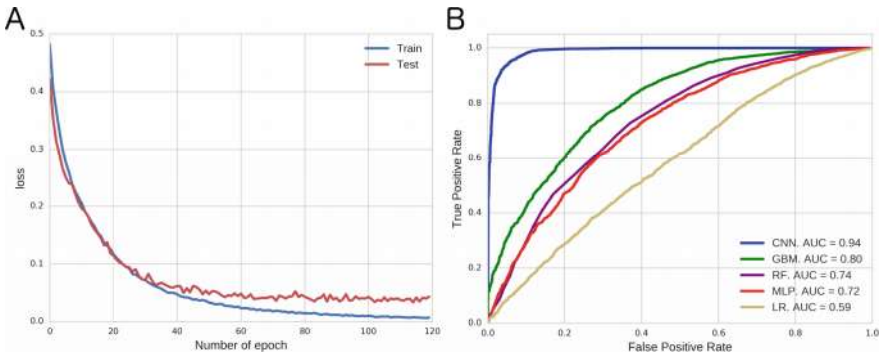

Fig. 3. Performance of best models. (A) Learning curve of the best CNN trained on spectrogramm features for "B" experiment. (B) ROC curves for classifiers trained on MFCC features for "B" experiment.

6 Conclusion

In this paper, we proposed CNN-based approach called WDCNN to detect wheezes in lung sound data. The results showed by the approach on sound signals presented as spectrograms and MFCC features were compared with conventional machine learning models, and WDCNN reaches state-of-the-art performance. Our method outperformed other approaches in the scope in terms of performance measure demonstrating stable 0.96 AUC score on the Internet dataset. Also, several experiments were conducted to prove that WDCNN is the robust method, insensitive to the lung sound shifting and external noise. Our method showed better performance in comparison with the baseline methods.

因为他的也是小数据集？？

还验证了算法的鲁棒性；

cnn的泛化性和可靠性很好；

Our findings showed that the generalization capability and reliability of the proposed CNN-based method is high enough to use it in real-world conditions. Thus, WDCNN is a well-suited and reliable method to use both in clinics and by any people at home.

References

1. Reichert, S., Raymond, G., Christian, B., Andrès, E.: Analysis of respiratory sounds: state of the art. Clin. Med. Circ. Respirat. Pulm. Med. **2**, 45–58 (2008)
2. Bahoura, M., Lu, X.: Separation of crackles from vesicular sounds using wavelet packet transform. In: 2006 IEEE International Conference on Acoustics Speed and Signal Processing Proceedings (2006)
3. Yann, L., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
4. Farabet, C., Couprie, C., Najman, L., LeCun, Y.: Learning hierarchical features for scene labeling. IEEE Trans. Pattern Anal. Mach. Intell. **35**(8), 1915–1929 (2013)
5. Tompson, J.J., Jain, A., LeCun, Y., Bregler, C.: Joint training of a convolutional network and a graphical model for human pose estimation. In: Advances in Neural Information Processing Systems, pp. 1799–1807 (2014)
6. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
7. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.: Natural language processing (almost) from scratch. J. Mach. Learn. Res. **12**(Aug), 2493–2537 (2011)
8. Hu, B., Lu, Z., Li, H., Chen, Q.: Convolutional neural network architectures for matching natural language sentences. In: Advances in Neural Information Processing Systems, pp. 2042–2050 (2014)
9. Bordes, A., Chopra, S., Weston, J.: Question answering with subgraph embeddings. arXiv preprint arXiv:1406.3676 (2014)
10. Palaz, D., Magimai-Doss, M., Collobert, R.: Analysis of CNN-based speech recognition system using raw speech as input. In: Proceedings of the 16th Annual Conference of International Speech Communication Association (Interspeech), pp. 11–15 (2015)
11. Mikolov, T., Deoras, A., Povey, D., Burget, L., Černocký, J.: Strategies for training large scale neural network language models. In: 2011 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), pp. 196–201 (2011)

这篇已经引用了；

12. Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.R., Jaitly, N., Kingsbury, B.: Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Sig. Process. Mag. **29**(6), 82–97 (2012)
13. Sainath, T.N., Mohamed, A.R., Kingsbury, B., Ramabhadran, B.: Deep convolutional neural networks for LVCSR. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8614–8618 (2013)
14. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems (2012)
15. Zhang, H., McLoughlin, I., Song, Y.: Robust sound event recognition using convolutional neural networks. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 559–563 (2015)
16. Zhang, J., Ser, W., Yu, J., Zhang, T.: A novel wheeze detection method for wearable monitoring systems. In: 2009 International Symposium on Intelligent Ubiquitous Computing and Education (2009)
17. Mayorga, P., Druzgalski, C., Morelos, R., Gonzalez, O., Vidales, J.: Acoustics based assessment of respiratory diseases using GMM classification. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology (2010)
18. Milicevic, M., Mazic, I., Bonkovic, M.: Classification accuracy comparison of asthmatic wheezing sounds recorded under ideal and real-world conditions. In: 15th International Conference on Artificial Intelligence, Knowledge Engineering and Databases (AIKED 2016), Venice (2016)
19. Bahoura, M., Pelletier, C.: Respiratory sounds classification using cepstral analysis and Gaussian mixture models. In: The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (2004)
20. Palaniappan, R., Sundaraj, K., Sundaraj, S.: A comparative study of the SVM and K-nn machine learning algorithms for the diagnosis of respiratory pathologies using pulmonary acoustic signals. BMC Bioinform. **15**, 223 (2014)
21. Wrigley, D.: Heart and Lung Sounds Reference Library. PESI HealthCare, Eau Claire (2011)
22. Shaharum, S., Sundaraj, K., Palaniappan, R.: A survey on automated wheeze detection systems for asthmatic patients. Bosnian J. Basic Med. Sci. **12**, 249 (2012)
23. Wei, H., Chan, C., Choy, C., Pun, P.: An efficient MFCC extraction method in speech recognition. In: Circuits and Systems (2006)
24. Ciresan, D.C., Meier, U., Masci, J., Gambardella, L.M., Schmidhuber, J.: High-performance neural networks for visual object classification. arXiv preprint arXiv:1102.0183 (2011)
25. Liaw, A., Wiener, M.: Classification and regression by randomForest. R News **2**(3), 18–22 (2002)
26. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. In: Annals of Statistics, pp. 1189–1232 (2001)

有很多用cnn做语音识别的工作！