Detection of Adventitious Respiratory Sounds based on Convolutional Neural Network

Renyu Liu
School of Electronic and Information Engineering
Soochow University
Suzhou, China
20185228013@stu.suda.edu.cn

Shengsheng Cai
Suzhou Melodicare Medical Technology Co., Ltd.
Suzhou, China
caishengsheng@mintti.cn

Kexin Zhang
College of Information Engineering
Liaoning University of Traditional Chinese Medicine
Shenyang, China
kxzh@sina.com

Nan Hu*
School of Electronic and Information Engineering
Soochow University
Suzhou, China
hunan@suda.edu.cn

***Abstract*—Nowadays, the respiratory disease has become one of the most dangerous diseases that threaten human health, especially in the developing countries. The early diagnosis of respiratory disease gives patients the opportunity to receive proper treatment in time, and hence artificial intelligent (AI) auscultation using electronic stethoscope may play a promising role here. The core idea of AI auscultation of respiratory disease is to detect or recognize two kinds of adventitious respiratory sounds related to respiratory diseases: wheeze and crackle. Constrained by the number of available data, in the traditional methods, subjectively defined features were extracted and used to detect these adventitious respiratory sounds. However, to make the detection results robust, the features had better to be learned automatically from the data, which can be realized by applying deep learning in a big data. In this paper, the convolutional neural network (CNN) is exploited to detect adventitious sounds. The data used in this study consists of two parts: the public database provided by the International Conference on Biomedical and Health Informatics (ICBHI) involving 126 subjects and our recorded pediatric auscultation data including 222 subjects. The detection performance of employed CNN is evaluated using ICBHI database, our pediatric auscultation database as well as the combination of them.**

Keywords—Adventitious respiratory sound; Deep learning; Convolutional neural network; Wheeze; Crackle

I. INTRODUCTION

The respiratory diseases, such as pneumonia, asthma, bronchitis, chronic obstructive pulmonary disease (COPD), etc., are very dangerous, if the correct diagnosis is not given in time. This issue is shown very serious in developing countries, due to the lack of medical resource. Every year, nearly 1 million children under 5 years old die of acute lower respiratory tract infections, the number of which exceeds the total number of deaths caused by HIV, malaria and tuberculosis in that age group [1]. At the clinics, the conventional ways to check for respiratory diseases include auscultation, bronchoscopy, chest CT scan, chest x ray and routine sputum culture, where the most simple and direct method is auscultation, in which a stethoscope is used to check whether the adventitious respiratory sounds exist. The adventitious sounds, which are related to respiratory diseases, adhere to the normal respiratory sound, and can be classified into the continuous sound (duration > 80 ms) and the discontinuous sound (duration < 25 ms) [2]. The most common continuous sound is wheeze, which is a musical, high-pitched sound, associated with airway narrowing or airflow limitation (e.g., asthma, COPD or tumor) [3]. The discontinuous sound almost equals the terminology “crackle”, including the nonmusical high-pitched fine crackle caused by explosive opening of small airways and the low-pitched coarse crackle caused by air bubble in bronchi or bronchiectatic segments, and is associated with various respiratory diseases such as pneumonia. Some adventitious sounds, e.g. fine crackle, may even be present before changes in radiology examination, which shows the value of auscultation. With the development of Internet of things (IoT), the emerging electronic stethoscopes make it possible to transcribe and analyze cardiopulmonary sounds automatically, which may save the decline of this art. Studying the artificial intelligent (AI) electronic stethoscope is fascinating, as we can examine the adventitious respiratory sounds ourselves, even at our home.

At present, there exist many methods on detecting adventitious respiratory sounds, by using traditional pattern recognition tools. In [4], the detection of crackle was realized by a morphological method, which detects the irregular eclipse image features in the short time Fourier transform (STFT)

This work was supported by Suzhou Science and Technology Project under Grant SYS2019029.
*Corresponding author

978-1-7281-3380-5/19/$31.00 ©2019 IEEE

spectrum of the received signal. In [5, 6], the detection of wheeze was given by using the entropy as feature. In [7], the feature obtained by fractional-order-based Hilbert transform was used to detect wheeze. A method for classifying continuous adventitious sounds, discontinuous adventitious sounds and normal respiratory sounds based on mutual information and weighted cepstrum was given in [8]. By wavelet analysis, crackles are detected based on the signal envelope at an optimal scale and classified based on the energy distribution in [9]. The SVM classifiers were often used to achieve the classification of adventitious respiratory sounds and normal respiratory sounds [10-12], where the features such as instantaneous frequency, Mel frequency cepstrum coefficient (MFCC) and spectral envelope were employed. The other clustering methods such as k-nearest neighbor (k-NN) method and Gaussian mixture model (GMM) are also used to detect adventitious sounds in [13-15]. Generally speaking, these methods extracted some subjectively chosen features for adventitious respiratory sounds, and then used an empirical threshold or pattern recognition method to determine the existence of these sounds. They can achieve good results, but were usually restricted to a small amount of data.

In practice, the adventitious sound detection methods will confront many severe tests: the recorded respiratory sound is probably polluted by ambient interference, clipping distortion, frictional noise cause by rubbing on the cloth, etc. The problems can be more serious for pediatric subjects, who are often uncooperative in medical examination. In this context, a reasonable way to design an adventitious sound detection method is to learn a deep neural network from a big database. In this paper, we study designing convolutional neural network (CNN) in detecting adventitious respiratory sound signals. Two databases are used, where the first one is the database including 126 subjects newly released by the International Conference on Biomedical and Health Informatics (ICBHI) [16], and the second one is our recorded database including 222 pediatric subjects. The two databases, especially the pediatric auscultation database, include many interference scenarios in practice. The procedures of data preprocessing and network training are given, and the detection performance is tested in every one of the databases as well as the mixture of them.

II. METHOD

A. Data Description

The database used in this study includes the public ICBHI data and our recorded pediatric auscultation data.

In the ICBHI database, 920 recordings were acquired from 126 subjects, recorded by two research teams in Portugal and Greece. Breathing experts had annotated crackle, wheeze, their combination or without pathological sound in each respiratory cycle. The subjects are ranged in all ages, and the recordings with durations from 10s to 90s were collected using heterogeneous equipments. Fig. 1 shows the STFT power spectra of recording examples containing crackle only, wheeze only, both of them and none of them.

In 2018, we acquired 508 recordings in pediatric clinics of several cooperative hospitals in Shenyang and Shanghai. These data are recorded from 222 pediatric subjects using the electronic stethoscope *Smartho-D2* (Fig. 2) developed by *minttihealth*. Six positions on the back of each subject were expected for data recording, while it is difficult to acquire all of them due to the complex situations in pediatric clinics such as subject uncooperation, loud crying, parents' whispering, etc. Fig. 3 shows a recording example confronting some of these issues. Hence, 1~6 recordings were acquired for each pediatric subject. Experienced respiratory physicians marked the styles and locations of adventitious sounds in each recording.

It is no doubt that judging whether the adventitious sounds exist in pediatric auscultation data is more difficult than that in data recorded from adults. One reason is that the average recording quality of pediatric data is worse than that of adult data, which can be partly observed from Fig. 1 and Fig. 3. Another reason is that the respiratory sounds of children are weaker than those of adults. It will lead to the different experimental results shown later in this paper.

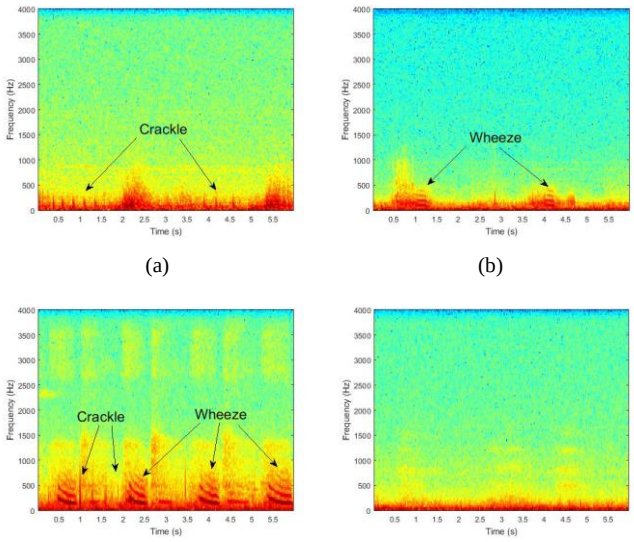


Fig. 1. The STFT power spectrum of recording examples from ICBHI: (a) crackle; (b) wheeze; (c) crackle and wheeze; (d): normal.

Fig. 2. The electronic stethoscope *Smartho-D2* used in pediatric auscultation

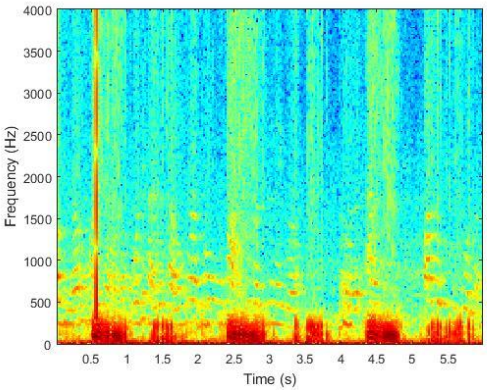


Fig. 3. The STFT power spectrum of a recording example in our recorded pediatric database, which is polluted by ambient noise and clipping distortion.

The recordings in the above mentioned two databases have various lengths and sampling rates, while the CNN used in this paper requires a fixed input size. Here we downsample the recordings into f_s = 8kHz, and divide them into multiple data segments. The length of each segment is determined as 2s, which is long enough to cover one or more adventitious sounds. Each segment is normalized, and the examples containing crackle and wheeze are shown in Fig. 4 and Fig. 5, respectively. The segments containing wheeze are much fewer than those containing crackle in both databases, and wheeze even sometimes appears in the same segment containing crackle. Hence, in this paper, the labels of segments are given binary: "with adventitious sounds" (1) or "without adventitious sounds" (0), according to the annotations of the two databases. In consideration of balance between each label, The ICBHI database are divided into 2086 segments where 1191 segments with *Label-1* and 895 segments with *Label-0*, and our pediatric database are divided into 1094 segments where 561 segments with *Label-1* and 533 segments with *Label-0*. In this paper, we focus on the segment-level detection of adventitious respiratory sounds, where the databases with relatively larger sizes can be used.

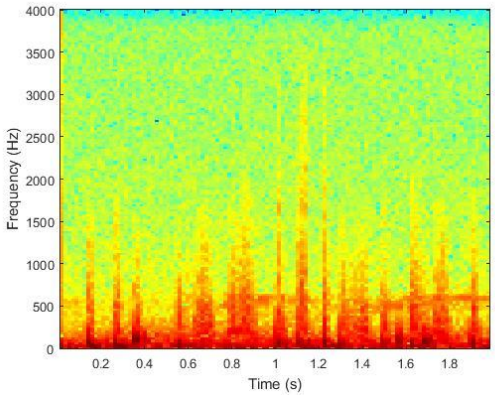


Fig. 4. The STFT power spectrum of a segment example containing crackle

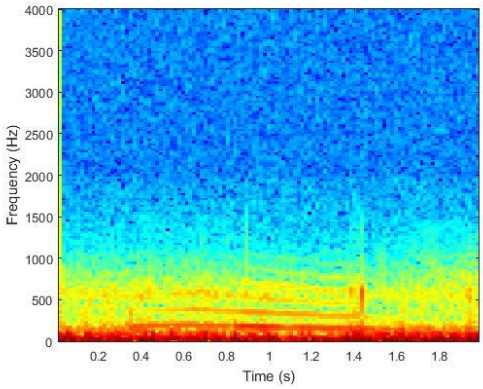


Fig. 5. The STFT power spectrum of a segment example containing wheeze

B. *Preprocessing*

The one-dimensional (1-D) recording segments are converted to a 3-D formulation before input to CNN. Here the Log *Mel*-filterbank (LMFB) of acoustic signal [17] is used. The procedure of LMFB transform is listed as follows.

Firstly, the STFT of each segment is calculated, where each segment is divided into M frames with $N_{\text{FFT}} = 1024$ length and 50% overlapping. Hence $M = 31$ frames are derived from each segment, where zero padding is used in the last frame. Denote $x_m(n), n = 0,1,...,N_{\text{FFT}} - 1$ as the samples in the *m*th frame for a certain segment. The fast Fourier transform (FFT) of $x_m(n)$ is calculated as $Y_m(k)$, which is given by

$$Y_m(k) = \sum_{n=0}^{N_{\text{FFT}}-1} x_m(n)h(n)e^{-j2\pi kn/N}, k = 0,1,...,N_{FFT}/2-1, \quad (1)$$

where $h(n)$ is the Hanning window.

Secondly, $|Y_m(k)|^2$ is filtered by a *Mel* filter bank. This *Mel* filter bank consists of Q linear-interval (50% overlapping) triangular filters $\Psi_q, q = 1,2,...,Q$ in *Mel* frequency domain, where the *Mel* frequency is displayed as

$$f_{\text{Mel}}(f) = 2959 \times \log_{10}(1 + f/700),\ f \sim [0,\ f_s/2]. \quad (2)$$

$|Y_m(k)|^2$ filtered by each *Mel* filter is given by

$$y_m(q) = \sum_{k=0}^{N_{FFT}/2} |Y_m(k)|^2 \Psi_q(k), q = 1,2,...,Q. \quad (3)$$

Lastly, the Q elements of LMFB of $x_m(n)$ are given by

$$F_m(q) = \log[y_m(q)], q = 1,2,...,Q. \quad (4)$$

Now we have obtained an LFMB matrix $\mathbf{F}$ with size $Q \times M$ for each segment, and then we transform it into 3 channels (displayed as Fig. 6). The first channel is given by $\mathbf{F}[:, 1:M-2]$. The second channel is calculated by the 1st-

order difference $\Delta_1 = \mathbf{F}[:, 2:M-1] - \mathbf{F}[:, 1:M-2]$. The third channel is obtained by the 2nd-order difference $\Delta_2 - \Delta_1$, where $\Delta_2 = \mathbf{F}[:, 3:M] - \mathbf{F}[:, 2:M-1]$. The second and third channels are used to exploit the dynamic features of the segment, while the LMFB can only represent the static features. To let the data easily handled by CNN, we make the input matrix in each channel be square, i.e. $Q = M - 2$.

The last step before applying CNN is to normalize the matrix in each channel, which is given by

$$\tilde{x} = (x - x_{\min}) / (x_{\max} - x_{\min}), \tag{5}$$

where $x_{\min}$ and $x_{\max}$ represent the minimum and maximum entries of the corresponding matrix.

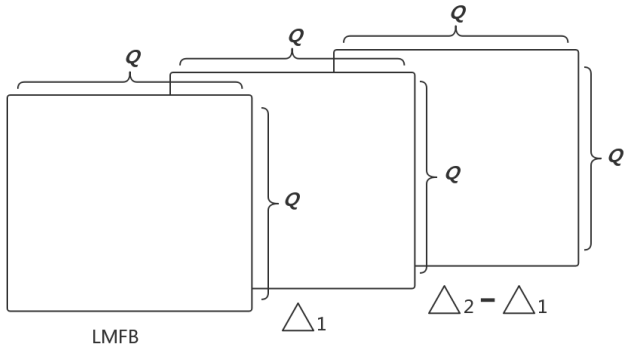


Fig. 6. The input channels for CNN

C. CNN

The CNN used in this paper is deployed as: after the input layer, two convolutional layers are followed by a pooling layer and this structure is repeated twice, and then a fully connected layer is given followed by the output layer. The architecture of the CNN employed is given in Fig. 7. The input layer corresponds to the three-channel formulation of the input data. In the convolutional layer, a suitable convolution kernel is employed (Fig. 8), and the actual output of convolutional layer is obtained by passing convolution calculation results through the *ReLU* function. The pooling layer is chosen as the maximum pooling, which extracts the most important features and is immune to location variation [18]. The activation function of the output layer uses the logistic *sigmoid* function, defined as $\sigma(z) = 1/[1 + \exp(-z)]$. The output results are mapped to a range in (0, 1), and can be deemed as the probability of adventitious respiratory sounds being detected in the input data segment.

The initialization of weights in this CNN is given by the *truncated_normal* function in *TensorFlow*, which generates weights according to a truncated normal distribution. The standard deviation of truncated normal distribution is set to be 0.1. The Adam optimizer is used in the training process and the learning rate is set to be 0.01. To prevent over-fitting, Dropout learning is added to the training of CNN, which makes part of the neurons randomly discarded. It means that when propagating forward, we let the activation value of a neuron stop working at a certain probability, hence making the model more extensive. The model will not rely too much on some local features, which may prevent the emergence of over-fitting. L_2 Regularization is also added to the forward propagation process of the network and the regularized coefficient is set to be 0.01, which can effectively prevent over-fitting.

III. Experimental Results

A. Evaluation of model

Each of the two databases is divided into a training set (75%) and a test set (25%). In the process of training, the training set was also randomly divided into two parts: 75% for training and 25% for validation. The purpose of this process is to carry out cross-validation, and then adjust the structure and parameters of the network by observing the accuracy of the validation set. The accuracy is calculated by

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}, \tag{6}$$

where *TP*, *TN*, *FP* and *FN* denote true positive, true negative, false positive and false negative, respectively.

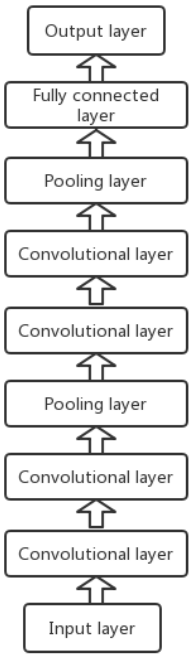


Fig. 7. The architecture of the CNN employed

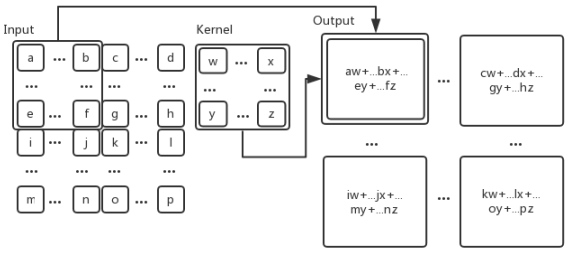
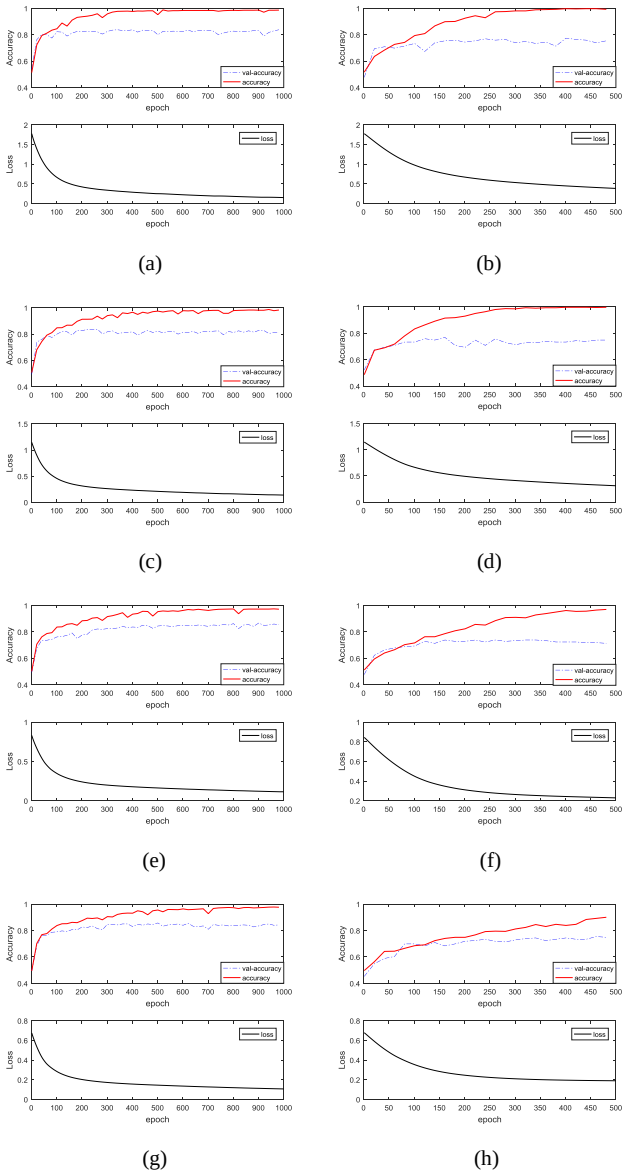


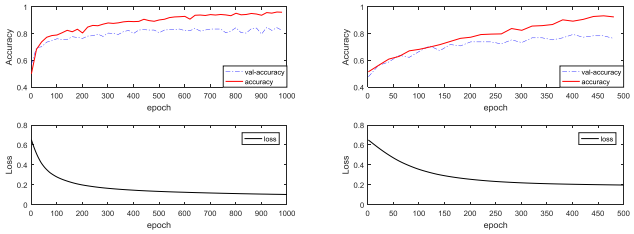
Fig. 8. Calculation process of convolutional layer

B. *Results*

When the CNN framework is determined, we need to set the size of the convolutional kernel in each convolutional layer. Convolutional kernels with different sizes will have different effects on the results. Here we restrict that the size of the convolutional kernels in the front convolutional layers is not smaller than that in the latter convolutional layers. In Fig. 9, the accuracy and model loss results for different sizes of convolutional kernels are plotted, where the legend of "val-accuracy" denotes the accuracy for the validation dataset and the legend of "accuracy" is the accuracy for the training dataset. (a), (c), (e), (g) and (i) of Fig. 9 are the training results by ICBHI database, and (b), (d), (f), (h) and (j) of Fig. 9 are the training results by our pediatric database. There are large gaps between the accuracy curve and the val-accuracy curve in (a), (b), (c), (d) and (f). These gaps in (e), (g), (h), (i) and (j) are smaller, and the convolutional kernels in (g) and (h) of Fig. 9 are favorable due to the stability and relative high accuracy.



(a) (b) (c) (d) (e) (f) (g) (h)



(i) (j)

Fig. 9. Accuracy and model loss with different sizes of convolutional kernels. (a) and (b): (5×5)-(5×5)-(5×5)-(5×5); (c) and (d): (5×5)-(5×5)-(5×5)-(3×3); (e) and (f): (5×5)-(5×5)-(3×3)-(3×3); (g) and (h): (5×5)-(3×3)-(3×3)-(3×3); (i) and (j): (3×3)-(3×3)-(3×3)-(3×3).

After getting the optimal convolutional kernel size: (5×5)-(3×3)-(3×3)-(3×3), we mix the ICBHI database with our pediatric database, and use this mixed database to train CNN with the same structure. Fig. 10 shows such a training result.

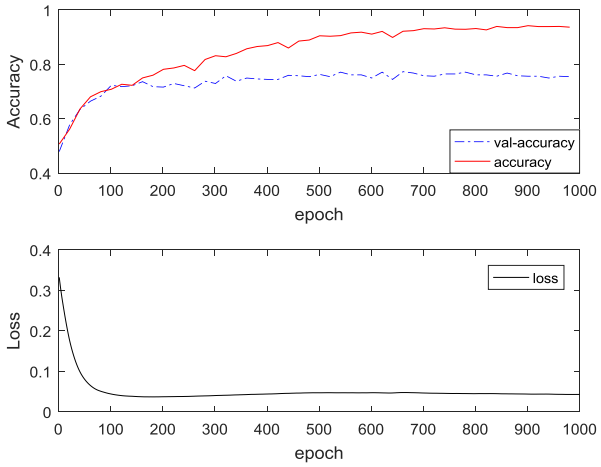


Fig. 10. Plot of accuracy and model loss of the Mixed database

Finally, we test the adventitious sound detection performance of the CNN model with the test sets of ICBHI database, our pediatric database and the mixture of them. Table 1 shows these results.

TABLE I. TEST RESULTS

	ICBHI database	Our pediatric database	Mixed databases
Accuracy	81.62%	69.72%	61.02%

C. *Discussion*

The CNN designed in this paper aims at segment-level detection of adventitious respiratory sounds. The recording-level detection can also be carried out by the following procedure: firstly multiple segments are derived by using overlapped sliding windows for each recording, secondly the segment-level detection is performed using CNN for each segment, and lastly the detection results of all the segments are combined to give the final recording-level detection.

From the experimental results, the accuracy of all cases starts at about 50% and grows rapidly. The gap between the training accuracy and the validation accuracy for our pediatric database is larger than that of ICBHI dataset, and the loss function is smaller than that of ICBHI database. In Fig. 10, the training accuracy for the mixed dataset can reach about 94% and the validation accuracy can reach about 76% after 1000 epochs, while accuracy given by its test dataset is only 61.02%. These results indicate that the method proposed in this paper is more suitable for ICBHI database. Compared to the ICBHI database including many adult subjects, our database obtained from pediatric auscultation has confronted more complex situations: the total number of efficient data being much less due to some subjects' uncooperation, the recording environment in pediatric clinics being much noisy, the energy of children's respiratory sound signals being much weaker, etc. Furthermore, as the detection accuracy of the test sets in the mixed database is much worse than that in the other two databases, it may indicate that there is a strong data variation between the two databases, and hence different detection methods may be applied for them.

IV. CONCLUSIONS

The usage of CNN in detecting adventitious respiratory sounds has been studied in this paper. ICBHI database as well as our recorded pediatric database was used for training and testing the CNN. The detection was performed on segment-level, and LMFB was used for preprocessing. By experimental results, the most suitable sizes of convolutional kernels of CNN were obtained, and the training accuracy, validation accuracy and testing accuracy were given for the two databases as well as the mixture of them. The method worked better for the ICBHI database, while for our pediatric database and the mixed database the detection accuracy still needs to be promoted. In the future work, we would expand our database as well as study the feature extraction and deep learning method, to achieve the better adventitious respiratory sound detection performance for patients in all ages.

REFERENCES

[1] L. Liu, S. Oza, D. Hogan, Y. Chu, J. Perin, J. Zhu, et al. "Global, regional, and national causes of under-5 mortality in 2000-15: an updated systematic analysis with implications for the sustainable development goals," The Lancet, vol. 388, 2016, pp. 3027-3035.

[2] R. Pramono, S. Bowyer, and E. Rodriguez, "Automatic adventitious respiratory sound analysis: A systematic review," PLOS ONE, vol. 12, May 2017.

[3] B. Abraham, G. Izbicki, and S. Kraman, "Fundamentals of lung auscultation," New England Journal of Medicine, vol. 370, Feb 2014, pp. 744-751.

[4] K. Zhang, X. Wang, F. Han, and H. Zhao, "The detection of crackles based on mathematical morphology in spectrogram analysis," Technology and Health Care, vol. 23, 2015, pp. S489-S494.

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

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

[7] Z. Li, and X. Wu, "Wheeze detection using fractional Hilbert transform in the time domain," IEEE Biomedical Circuits and Systems Conference (BioCAS), 2012, pp. 316–319.

[8] F. Jin, F. Sattar, and S. Krishnan, "Log-frequency spectrogram for respiratory sound monitoring," IEEE International Conference on Acoustics, speech and signal processing (ICASSP), 2012, pp. 597–600.

[9] M. Du, F. Chan, F. Lam, and J. Shun, "Crackle detection and classification based on matched wavelet analysis," International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), 1997, pp. 1638–1641.

[10] D. Chamberlain, R. Kodgule, D. Ganelin, V. Miglani, and R. Fletcher, "Application of semi-supervised deep learning to lung sound analysis," 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016, pp. 804-807.

[11] I. Mazić, M. Bonković, and B. DzĬaja, "Two-level coarse-to-fine classification algorithm for asthma wheezing recognition in children's respiratory sounds," Biomedical Signal Processing and Control, vol. 21, 2015, pp.105-118.

[12] M. Wiśniewski, and T. Zieliński, "Joint application of audio spectral envelope and tonality index in an e-asthma monitoring system," IEEE Journal Biomedical and Health Informatics, vol. 19, 2015, pp. 1009-1018.

[13] B. Mohammed, "Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes," Computers in Biology and Medicine, vol. 39, 2009, pp. 824-843.

[14] P. Mayorga, "Acoustics based assessment of respiratory diseases using GMM classification," Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2010, pp. 6312-6316.

[15] S. Alsmadi, and Y. Kahya, "Design of a DSP-based instrument for real-time classification of pulmonary sounds," Computers in Biology and Medicine, vol. 38, 2008, pp. 53-61.

[16] B. Rocha, D. Filos, L. Mendes, I. Vogiatzis, E. Perantoin. E. Kaimakamis, et al. "A respiratory sound database for the development of automated classification," International Conference on Biomedical and Health Informatics,Springer, Singapore, vol. 66, 2017.

[17] Y. Jung, Y. Kim, H. Lim, and H. Kim, "Linear-scale filterbank for deep neural network-based voice activity detection," 20th Conference of the Oriental Chapter of the International Coordinating Committee on Speech Databases and Speech I/O Systems and Assessment, 2017, pp. 1-5.

[18] M. Islam, B. Siddique, S. Rahman, and T. Jabid, "Image recognition with deep learning," International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), IEEE Computer Society, 2018.