IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 1, 2019, accepted March 4, 2019, date of publication March 8, 2019, date of current version March 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2903859

Triple-Classification of Respiratory Sounds Using Optimized S-Transform and Deep Residual Networks

HAI CHEN[1,2], (Member, IEEE), XIAOCHEN YUAN[1], (Member, IEEE), ZHIYUAN PEI[2], MIANJIE LI[1], AND JIANQING LI[1], (Senior Member, IEEE)

[1]Faculty of Information Technology, Macau University of Science and Technology, Macau 999078, China
[2]School of Information Technology, Beijing Normal University, Zhuhai 519086, China

Corresponding author: Xiaochen Yuan (xcyuan@must.edu.mo)

ABSTRACT Digital respiratory sounds provide valuable information for telemedicine and smart diagnosis in an non-invasive way of pathological detection. As the typical continuous abnormal respiratory sound, wheeze is clinically correlated with asthma or chronic obstructive lung diseases. Meanwhile, the discontinuous adventitious crackle is clinically correlated with pneumonia, bronchitis, and so on. The detection and classification of both attract many studies for decades. However, due to the contained artifacts and constrained feature extraction methods, the reliability and accuracy of the classification of wheeze, crackle, and normal sounds need significant improvement. In this paper, we propose a novel method for the identification of wheeze, crackle, and normal sounds using the *optimized S-transform* (OST) and *deep residual networks* (ResNets). First, the raw respiratory sound is processed by the proposed OST. Then, the spectrogram of OST is rescaled for the Resnet. After the feature learning and classification are fulfilled by the ResNet, the classes of respiratory sounds are recognized. Because the proposed OST highlights the features of wheeze, crackle, and respiratory sounds, and the deep residual learning generates discriminative features for better recognition, this proposed method provides reliable access for respiratory disease-related telemedicine and E-health diagnosis. The experimental results show that the proposed OST and ResNet is excellent for the multi-classification of respiratory sounds with the *accuracy*, *sensitivity*, *and specificity* up to 98.79%, 96.27%, and 100%, respectively. The comparison results of the triple-classification of respiratory sounds indicate that the proposed method outperforms the deep-learning-based ensembling convolutional neural network (CNN) by 3.23% and the empirical mode decomposition-based artificial neural network (ANN) by 4.63%, respectively.

INDEX TERMS Deep residual networks (ResNet), optimized S-transform (OST), respiratory sounds classification, crackle and wheeze detection.

I. INTRODUCTION

Digital respiratory sounds provide important clinical characteristics of normal and pathological index, which offer the crucial basis for telemedicine and smart diagnosis. Instead of relying on the professional experience, computerized analysis of respiratory sounds offers objective detection and diagnosis of adventitious pathological sounds. Moreover, the automatic detection of adventitious respiratory sounds facilitates the long-term monitoring and treatment in an economical and convenient way. Among them, wheeze and crackle are typical continuous and discontinuous adventitious respiratory sound respectively which have been studied for decades. As a musical high-pitched continuous sound with the typical frequency from 100 Hz to 5000 Hz and the duration above 80 msec [1], wheeze is an important symptom of asthma and Chronic Obstructive Pulmonary Disease (COPD). As compared with wheezes, the nonmusical and short discontinuous crackles include fine crackles and coarse crackles. The frequency of the fine crackles is about 650 Hz and the duration is about 5 msec, while the frequency of the coarse crackles is about 350 Hz and the duration is about 15 msec. Crackles is

The associate editor coordinating the review of this manuscript and approving it for publication was Carlo Cattani.

typically associated with obstructive lung diseases including COPD, chronic bronchitis, pneumonia, and lung fibrosis [2].

Traditionally, the three steps needed to detect or classify wheeze, crackle and normal sounds are pre-processing, feature extraction and classification [3]. In pre-processing stage, the artifacts including con hearts sounds, background noises and contact interference are removed by filters, such as the band-pass filter of Butterworth [4] or adaptive filters [5]. Next, feature extraction is performed mainly based on spectral features [6], [7], eigen value of singular spectrum analysis (SSA) [8], Mel-frequency cepstral coefficients (MFCCs) [9], ensemble empirical mode decomposition [10], wavelet based musical features [11], statistical features of S-transform [12], local binary patterns (LBP) features [13], energy envelope [14], entropy-based features [15] or combination features among the above mentioned methods. In the last classification stage, conventional methods is based on the empirical threshold [16], [17]. Recently, great improvements have achieved with the machine learning launch. The widely used machine learning based methods for the classification of respiratory sounds include Gaussian mixed model (GMM) [18], support vector machine (SVM) [15], k-nearest neighbors (KNN) [19], extreme learning machine (ELM) [20], logistic regression method (LRM) [7]. In [21], it is shown that although the methods above mentioned are reliable for bi-classification of wheeze or crackle and normal sounds, the triple-classification of wheeze, crackle and normal sounds is still challenging. The handcrafted features and single classifier cannot adapt well for the complex recognition among wheeze, crackle and normal sounds. Therefore, some breakthroughs are needed for the multi-classification of the respiratory sounds by some comprehensive feature extraction. With the development of deep neural networks, the other way to achieve multi-classification emerged recently. Instead of extracting features of respiratory sounds with statistical or handcrafted [22] methods, adapted-well features can be generated by deep learning networks. Considering the effectiveness of deep learning in classifying images [23], we propose to preprocess the respiratory sounds before the deep learning-based methods can be employed. The preprocessing step is to transform the respiratory sound into the corresponding feature maps in time-frequency domain or extracting feature coefficients with the methods of STFT, MFCC, ST and so on. The classification is implemented by training and testing with the methods of GMM, ANN, CNN and ensembled CNN. In order to improve the accuracy of the multi-classification for respiratory sounds, the networks become deeper and deeper. Apart from disadvantages of long training-time and hardware configuration requirements, the bottleneck of deep-networks is the vanishing gradient when the networks become deep.

This paper presents a novel respiratory classification method based on ***Optimized S-transform (OST) and deep residual networks (ResNet)*** to recognize wheeze, crackle and normal sounds. The proposed method combines the merits of S-transform and deep-learning networks because the OST highlights the features of respiratory sounds in the preprocessing stage and the ResNet solve the vanishing gradient problem for deep-networks. Thus, the ResNet ensures the benefit of better performance with deeper networks. In the proposed method, the raw respiratory sounds is performed the OST without removing heart sounds and other artifacts. Then, the OST spectrogram of a breathing segment, including inspiratory and expiratory phases, is rescaled as input of the proposed deep ResNet, thus fulfilling the classification. In this study, the ResNet is trained and tested with a challenging dataset which includes records of noisy respiratory sounds and can evaluate the validation of proposed method. Experimental results show the superiority of the proposed OST and deep ResNet based method with the triple-classification accuracy of 98.79%, the sensitivity of 96.27% and specificity of 100% respectively. The organization of this paper is as follows. Section II discusses the proposed method. Section III describes the experimental data, experimental results, discussions and comparisons. Finally, conclusions are drawn in Section IV.

II. PROPOSED METHOD

In this study, we propose a novel method of ***deep ResNet*** fed by rescaled feature maps of the proposed ***OST*** to classify wheezes, crackles and normal sounds. The flowchart of the proposed OST and ResNet-based triple-classification is shown in Figure 1. In detail, the raw segment of respiratory sound is firstly performed the OST, then the OST coefficients are described as the corresponding spectrogram and rescaled into three fixed-size feature maps of the RGB values versa the rescaling processing. After the processing of training and classification with ResNet, wheeze, crackle and normal are recognized. More details are depicted as follows. In Section A, the OST is introduced and then the rescaling-preprocessing is briefly described in section B. Lastly, the training and classification with Resnet are described in section C.

A. PROPOSED OPTIMIZED S-TRANSFORM (OST)

In methods of deep-learning based classification, feature maps of respiratory sounds are fed into the classifier. In [22] and [24], spectrogram images of respiratory segments are the input of CNN. Although the STFT based spectrogram-image includes the characteristics of respiratory sounds, some mixed artifacts cannot be filtered are also included in this spectrogram, which hinder the feature learning with networks and the performance of classifiers. In our proposed OST based ResNet method, OST is performed for feature maps of respiratory segments. In time-frequency domain, compared with fixed window based STFT, S-transform [25] highlights frequency related features with the special frequency-dependent window, which varies wider at low frequencies and narrower at high frequencies. The S-transform is donated as follows

$$St_{seg}(\tau, f) = \int_{-\infty}^{+\infty} seg(t)\omega(t-\tau, f)e^{-2\pi jft}dt \qquad (1)$$

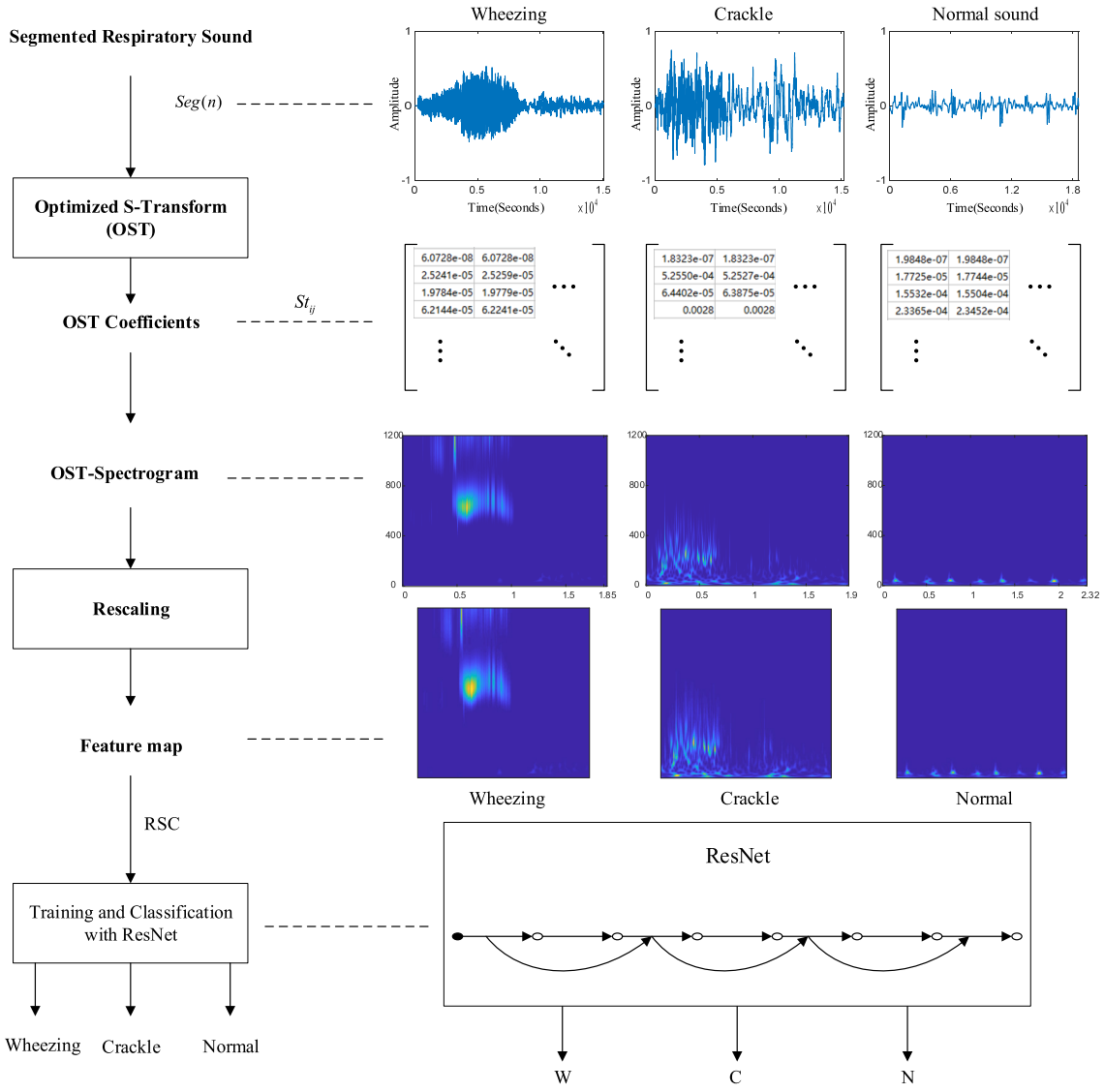

FIGURE 1. Flowchart of proposed ResNet-based respiratory sounds classification.

where $seg(t)$ is respiratory signal, $\omega(t-\tau, f)$ is the frequency-depended window function and τ is the time-variable.

For adventitious respiratory sounds, S-transform displays the better time-frequency resolution due to the width-varying window which makes the display clearly both at low frequencies and high frequencies. Apart from that, the S-transform uniquely offers the absolutely referenced local phase information and correspondent frequency features simultaneously, which can be described from the discrete of S-transform as

$$St_{seg}[m,n] = \begin{cases} \sum_{k=0}^{N-1} Seg[n+k] \cdot e^{-2\pi^2 k^2/n^2} \cdot e^{2\pi kmj/N} & (n \neq 0) \\ \frac{1}{N} \sum_{k=0}^{N-1} Seg[k] & (n = 0) \end{cases} \quad (2)$$

where, $m, n \in \{0, 1, \ldots, N-1\}$ indicate the row and column of the coefficient of S-transform matrix respectively, $Seg(n)$ is the discrete form of the respiratory segment, and N is the number of the segment samples.

According to the definitions of wheeze and crackle, which are examples of the classic continuous and discontinues adventitious respiratory sounds, the key to differentiate them from other respiratory sounds is the index of frequency and the corresponding duration, such as 100ms above for wheeze and less than 15ms for crackles respectively. The S-transform is of these merits and the illustrations of transform performance with respiratory sounds via STFT and ST can be found in Figure 2.

As Figure 2 shows, the frequency-depended window of S-transform allows the swiftly track of signal changes in frequency, phases and amplitude. However, improvements can be done because the frequency range between

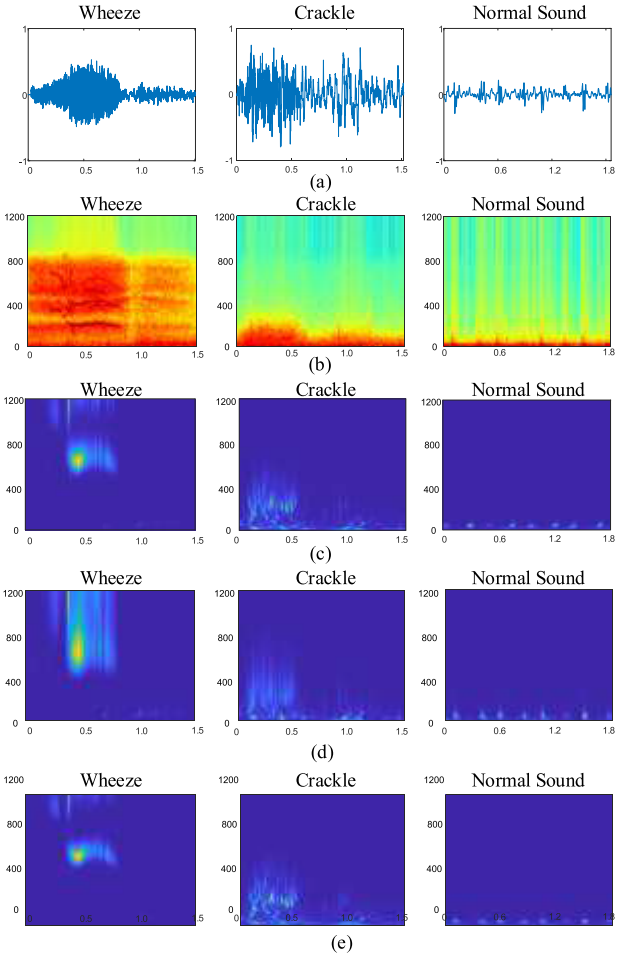

FIGURE 2. Transform spectrograms of STFT, ST and OST for wheeze, crackle and normal sound. (a) Respiratory segments with wheeze, crackle and normal sound. (b) STFT of the wheeze, crackle and normal sound. (c) ST of the wheeze, crackle and normal sound. (d) OST of the wheeze, crackle and normal sound with the parameter $\beta < 1$. (e) OST of the wheeze, crackle and normal sound with the parameter $\beta > 1$.

100 Hz to 2000 Hz deserves to be displayed more detailly due to the features of respiratory sounds. To fulfill the more precisely control of the window-variation, generalized S-transform [26] is employed and defined as

$$OSt_{seg}(\tau, f) = \int_{-\infty}^{+\infty} seg(t)\omega(t-\tau, \sigma(f))e^{-2\pi jft}dt \quad (3)$$

where, $\sigma(f) = \frac{\beta}{|f|}$. With the variation of the parameter β, the optimized S-transform can finely control the variation speed of the window-width with frequency-variation, the coefficients of S-transform are displayed and highlighted focus on the special frequency range pointed for the respiratory sounds. The illustration performance of generalized S-transform with respiratory sounds of wheeze, crackle and normal sound is shown in Figure 2. The performances of generalized S-transform with different values of the parameter β can be found in Figure 2 (d) and (e) respectively. Because the transform coefficients are used for the ResNet,

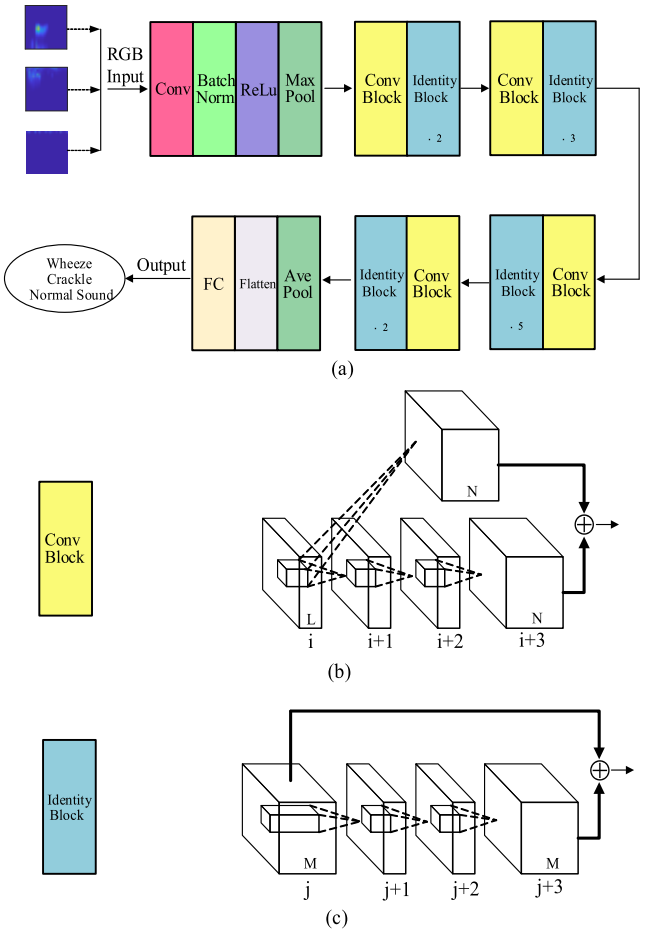

FIGURE 3. Structure of ResNet 50 for respiratory sounds classification. (a) Structure of ResNet with layers 50. (b) Conv block. (c) Identity block.

the optimization is fulfilled by the combination with the ResNet performances.

B. RESCALING THE SPECTROGRAM OF OST

Although the coefficients matrix of the S-transform represents the characteristics of the respiratory sounds in frequency, phases and amplitude, it is necessary to visually display its corresponding spectrogram [27] with energy distribution in time-frequency plane by Eq. 4 for the preprocessing of the next step.

$$\left|St_{seg}(\tau, f)\right|^2 = \left|\int_{-\infty}^{+\infty} seg(t)\omega(t-\tau, f)e^{-2\pi jft}dt\right|^2 \quad (4)$$

where St_{seg} is the coefficients of generalized S-transform. As shown in Figure 1, the OST spectrogram of the respiratory segment is rescaled into three fixed size 224 × 224 RGB images [28], [29] by bilinear method, and fed to the deep learning based classifiers as input of feature maps.

C. TRAINING AND CLASSIFICATION USING RESNET

Deep-learning based classifiers improve accuracies in images multi-classification greatly. In our study, we propose ***ResNet***

with OST based feature map to classify wheeze, crackle and normal sound. Figure 3 shows the flowchart of training and classification with Resnet. In detail, three RGB-maps of the rescaled feature map is fed into ResNet with layers 50 [30] due to the balance between the depth and performance. As Figure 3 (a) shown, the input feature map is passed through three steps of the ResNet structure and finally the class of wheeze, crackle or normal is output.

STEP-1 (ResNet Preprocessing): Orderly, the three RGB images of an input feature map are preprocessed and normalized with convolution, batch-normalization, ReLU [31] and Maxpooling.

STEP-2 (Feature Maps PROCESSING With Residual Learning): Conv Blocks and Identity Blocks are orderly included in this structure as Figure 3 shown. Where, $IdentityBlock \times 2$ means two consecutive Identity Blocks are stacked together, and so on. Detailly, there are two kinds of layers-shortcuts in residual blocks namely Conv Block and Identity Block in the ResNet structure. The structures of them are described in Figure 3 (b) and (c) respectively. In a residual block, when increase dimensions occurs between the two shortcut-layers, such as dimensions $L < N$, the 1×1 convolutions are taken to match their dimensions of lay i and lay $i + 3$ before the sum of them are fed to lay $i + 4$ and the residual block is defined as the Conv Block. On the other hand, when the two shortcut-layers are of the same dimensions, a direct shortcut namely identity is employed with lay j and lay $j + 3$ to extra add the output of lay j to lay $j + 4$, which is defined as Identity Block.

STEP-3 (Classification Implementation): With step-2, the output processed features $7 \times 7 \times 2048$, are average-pooled to $1 \times 1 \times 2048$ and flattened to 2048, after the processing of full connection (FC) and the softmax activation, wheeze, crackle and normal can be classified.

III. EXPERIMENTAL RESULTS

In this section, dataset of our study, training and classification based on the ResNet with feature maps of STFT, ST and the OST are depicted and discussed respectively. Experiments are carried out with the NVIDIA Titan V of GV100 GPU.

A. DATASET AND EVALUATION MATRIX

The dataset in this study are composed of three kinds of recordings of noisy respiratory sounds of Int. Conf. on Biomedical Health Informatics (ICBHI) Scientific challenge database [32]. The recordings include wheezes, crackles and normal sounds, which are recorded by the digital stethoscopes of 3M littleman3200 and WelchAllyn Elite (Meditron). In total, the 489 recordings include 44 records of wheezes, 136 records of crackles and 309 records of normal sounds, the corresponding segments of breath cycles include 149 breath cycles of wheezes, 386 breath cycles of crackles and 1125 breath cycles of normal sounds respectively. The dataset is randomly departed into two sub-datasets, and one sub-dataset including 70% data of our dataset is used for training and the other sub-dataset including the left 30% data of our dataset is used for testing. The training and testing are performed by the proposed ResNet with 50 layers and the rescaled feature maps are the input, described as section II. Our experimental results are evaluated by the index of Accuracy (%), Sensitivity (SE) (%) [33] and Specificity (SP) (%) at segment level and the definitions are given in (5), (6) and (7) as follows.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$SE = \frac{TCA}{TNA} \tag{6}$$

$$SP = \frac{NCN}{TNN} \tag{7}$$

where, TP means true positive, TN means true negative, FP means false positive and FP means false positive respectively. *TCA* refers to the True Classified Abnormal segments, which means the number of correctly classified segments of wheezes and crackles, *TNA* refers to Total Number of Abnormal segments, wheezes and crackles, under the test. *NCN* refers to the Number of Correctly classified Normal segments and *TNN* refers to Total Number of Normal segments in the test. The experimental results are described in section B.

B. EVALUATION OF THE PROPOSED METHOD

In order to evaluate the effectiveness of the proposed OST and ResNet for the triple-classification of respiratory sounds, the three rescaled feature maps of STFT, ST and OST are applied to the ResNet-50 with different batch sizes and iterations respectively. The results are listed in Table 1 to Table 3 in the term of classification accuracy, sensitivity and specificity respectively and confirm the superiority and reliability of the proposed OST and ResNet for the recognitions of wheezes, crackle and normal sounds.

TABLE 1. Classification accuracy (%) using ResNet and STFT.

Iterations / Batch size	16	32	64
5000	93.57	90.16	85.14
10000	90.56	90.36	86.55
15000	**93.98**	89.96	87.35
20000	92.97	90.36	86.35

TABLE 2. Classification accuracy (%) using ResNet and ST.

Iterations/ Batch size	16	32	64
5000	97.19	97.59	95.38
10000	**97.79**	96.99	97.19
15000	97.19	97.39	96.79
20000	97.59	97.39	95.98

Table 1 shows the classification results of wheezes, crackle and normal sounds employing the ResNet 50 with different batch sizes and iterations using the STFT based feature

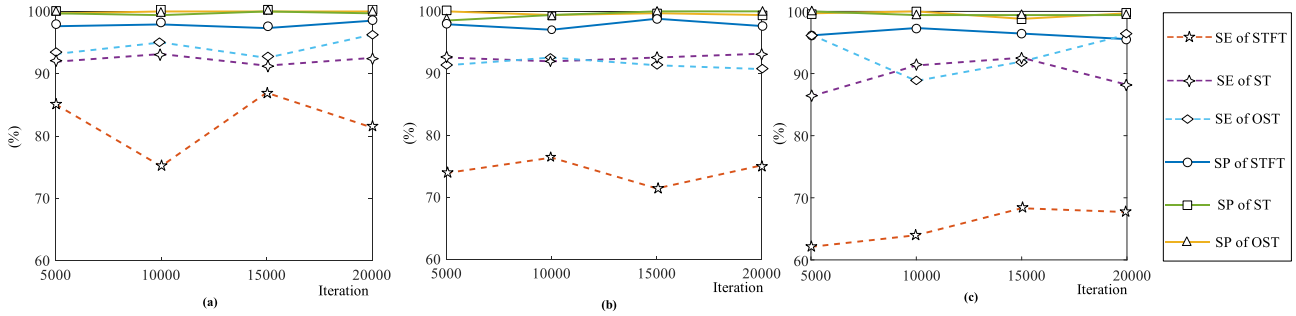

FIGURE 4. Classification results OF wheezes, crackles and normal sounds using ResNet with STFT, ST and OST. (a) ResNet with batch size 16. (b) ResNet with batch size 32. (c) ResNet with batch size 64.

TABLE 3. Classification accuracy (%) using ResNet and OST.

Iterations/ Batch size	16	32	64
5000	97.59	96.18	96.79
10000	97.99	97.19	95.98
15000	97.59	97.39	96.99
20000	**98.79**	97.19	98.39

TABLE 4. Performance comparison between different methods for the classification of wheezes, crackles and normal sounds.

Method	SE (%)	SP (%)	Accuracy (%)
MFCC - ANN	90	93.33	92.05
SSA- ANN	86.31	91.11	90
WT - ANN	86.66	92.96	91.66
EMD – ANN [31]	100	93.75	94.16
MFCC - KNN	N/A	N/A	79.94
MFCC - GMM	N/A	N/A	86.68
MFCC - SVM	83.63	99.04	91.12
LBP - KNN	N/A	N/A	67.28
LBP - GMM	N/A	N/A	69.07
LBP - SVM	68.33	77.27	71.21
MFCC - CNN	N/A	N/A	91.67
Ensembling CNN [30]	95.01	96.15	95.56
STFT - *ResNet*	**86.96**	**97.33**	**93.98**
ST - *ResNet*	**93.17**	**100**	**97.79**
Proposed OST - ResNet	**96.27**	**100**	**98.79**

N/A: Not mentioned in the reference papers.

maps as input. The best performance of classification accuracy is 93.98% with the batch size 16 and iteration 15000. In Table 2, with the batch sizes from 16 to 64 and iterations from 5000 to 20000 simultaneously, the best performance of the ResNet based on the ST feature maps is with classification accuracy 97.79% when the batch size is 16 and iterations are 10000 times. Compared with STFT and ST in Table 1 and Table 2 respectively, the proposed method of the ResNet based on OST feature maps is outperformed in triple-classifications of wheeze, crackle and normal sounds as shown in table 3 with accuracy 98.79% at the batch size 16 and iterations 20000.

In addition to this, the sensitivities and specificities of the ResNet based methods for the recognition of respiratory sounds are also shown in Figure 4. Where the batch sizes of ResNet are set as 16, 32 and 64 in Figure 4(a), (b) and (c) respectively with the iterations are 5000,10000,15000 and 20000 simultaneously. The best performance of ResNet with STFT feature maps is SE 86.96% and SP 97.33% respectively at the batch size 16 and iterations 15000 at Figure 4 (a). And the best performance of ResNet with the ST feature maps is the SE of 93.17% and the SP of 100% respectively at batch size 16, iteration 10000 at Figure 4(a). It is shown at Figure 4(a), (b) and (c) that the proposed method of the ResNet with OST outperforms with the STFT and ST. Among them, the best performance of proposed ResNet with OST is the SE of 96.27%, the SP of 100%, which is implemented under the batch size of 64 and iterations of 5000 at Figure 4(c).

Furthermore, the corresponding detailed recognition results of their best performance based on the ResNet with STFT, ST and OST are given at Figure 5. Detailly, ResNet with ST outperformed with STFT for wheezes, crackles and normal sounds recognition. And the ResNet with OST performs the best among them. In Figure 5 (c), 337 segments of the normal sounds are all recognized, 44 segments from the 45 testing segments of wheezes are recognized and 111 crackles are recognized from the 115 crackles respectively. The outperformance of the ResNet offer a reliable method for triple-classification of wheezes, crackles and normal sounds.

C. COMPARISON AND DISCUSSION

The proposed triple-classification of respiratory sounds method, ResNet with OST based feature map is compared with the recent methods as listed in table 4.

In [24], the MFCC and LBP based features are extracted respectively and the machine learning based classification models of popular KNN, SVM, GMM and CNN are

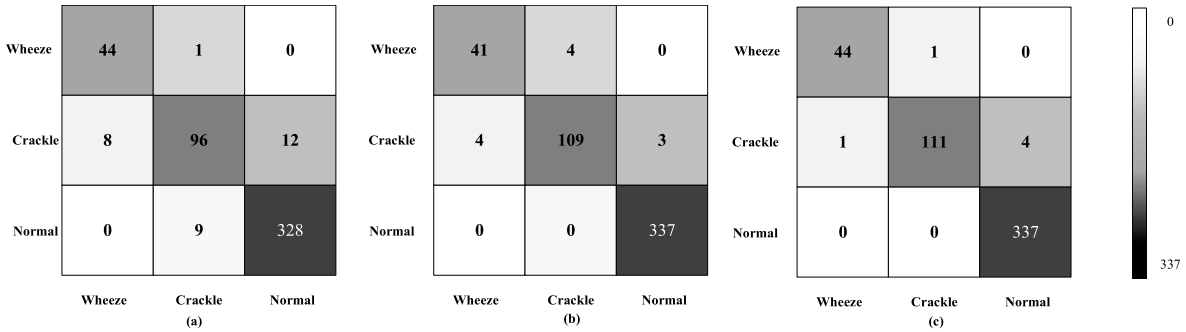

FIGURE 5. Confusion matrixes for (a) the ResNet and STFT classification of wheeze, crackle and normal sounds, (b) the ResNet and ST classification of wheeze, crackle and normal sounds, (c) the ResNet and OST classification of wheeze, crackle and normal sounds.

employed with the extracted features. As table 4 shows, the best performance is the accuracy of 91.67% with MFCC and CNN and with the proposed ensemble model the accuracy is achieved 95.56% in their test. In [34], the classification methods of ANN based on the features of MFCC, SSA and WT are employed respectively. The experimental results show the best performance is the accuracy of 94.16% with EMD and ANN. In our study, the deep ResNet with feature maps based on STFT, ST and OST are employed respectively. As table 4 shows, the deep ResNet based triple-classification method overperformed all the methods described in [24] and [34] with the ST features at the accuracy of 93.98%, SE of 96.27%, SP of 100% and with the OST features at the accuracy of 98.79%, SE of 96.27% and SP of 100% respectively. The proposed method of the ResNet with OST shows the superiority and reliability in triple-classification of the respiratory sounds.

IV. CONCLUSION

This paper proposes a deep learning method of the ResNet with OST to recognize wheezes, crackles and normal sounds. The proposed ResNet based method is implemented with the inputs of different rescaled spectrum maps, the STFT, ST and OST respectively. With the challenging dataset of noisy respiratory sounds, the experimental results show the outperformance of the proposed ***the ResNet with OST*** with the classification accuracy of 98.79%. Also, with the SE of 96.27% and the SP of 100%, the proposed method is reliable for the recognition of wheezes, crackles and normal sounds from respiratory sounds. In summary, the proposed ResNet using OST provides a reliable, convenient, and economical telemedicine diagnosis of respiratory diseases. The future work will focus on other deep-learning based methods for multi-classification and real-time diagnosis of pulmonary diseases.

ACKNOWLEDGMENT

The authors of this study sincerely thank M.D. Xiaobin Zheng for his cooperation in the analysis of pulmonary diseases.

REFERENCES

[1] A. Bohadana, G. Izbicki, and S. S. Kraman, "Fundamentals of lung auscultation," *New England J. Med.*, vol. 370, no. 8, pp. 744–751, Feb. 2014.

[2] R. X. A. Pramono, S. Bowyer, and E. Rodriguez-Villegas, "Automatic adventitious respiratory sound analysis: A systematic review," *PLoS ONE*, vol. 12, no. 5, May 2017, Art. no. e0177926.

[3] R. Palaniappan, K. Sundaraj, and N. U. Ahamed, "Machine learning in lung sound analysis: A systematic review," *Biocybern. Biomed. Eng.*, vol. 33, no. 3, pp. 129–135, 2013.

[4] C. Herley and M. Vetterli, "Wavelets and recursive filter banks," *IEEE Trans. Signal Process.*, vol. 41, no. 8, pp. 2536–2556, Aug. 1993.

[5] J. Gnitecki and Z. M. Moussavi, "Separating heart sounds from lung sounds," *IEEE Eng. Med. Biol. Mag.*, vol. 26, no. 1, p. 20, Jun. 2007.

[6] S. O. Maruf, M. U. Azhar, S. G. Khawaja, and M. U. Akram, "Crackle separation and classification from normal respiratory sounds using Gaussian mixture model," in *Proc. ICIIS*, New York, NY, USA, Dec. 2015, pp. 267–271.

[7] P. Bokov, B. Mahut, P. Flaud, and C. Delclaux, "Wheezing recognition algorithm using recordings of respiratory sounds at the mouth in a pediatric population," *Comput. Biol. Med.*, vol. 70, pp. 40–50, Mar. 2016.

[8] S. Içer and S. Gengeç, "Classification and analysis of non-stationary characteristics of crackle and rhonchus lung adventitious sounds," *Digit. Signal Process.*, vol. 28, pp. 18–27, May 2014.

[9] B.-S. Lin and B.-S. Lin, "Automatic wheezing detection using speech recognition technique," *J. Med. Biol. Eng.*, vol. 36, no. 4, pp. 545–554, Aug. 2016.

[10] M. Lozano, J. A. Fiz, and R. Jané, "Automatic differentiation of normal and continuous adventitious respiratory sounds using ensemble empirical mode decomposition and instantaneous frequency," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 2, pp. 486–497, Mar. 2016.

[11] M. Bahoura, "Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes," *Comput. Biol. Med.*, vol. 39, no. 9, pp. 824–843, 2009.

[12] R. Palaniappan, K. Sundaraj, S. Sundaraj, N. Huliraj, and S. S. Revadi, "A telemedicine tool to detect pulmonary pathology using computerized pulmonary acoustic signal analysis," *Appl. Soft Comput.*, vol. 37, pp. 952–959, Dec. 2015.

[13] M. Hassaballah, H. A. Alshazly, and A. A. Ali, "Ear recognition using local binary patterns: A comparative experimental study," *Expert Syst. Appl.*, vol. 118, pp. 182–200, Mar. 2019.

[14] Y. P. Kahya, E. C. Güler, B. Sankur, and T. Engin, "Detection and clustering analysis of crackles in respiratory sounds," in *Proc. 14th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 6, Paris, France, Oct./Nov. 1992, pp. 2527–2528.

[15] A. McCallum, D. Freitag, and F. C. N. Pereira, "Maximum entropy Markov models for information extraction and segmentation," in *Proc. ICML*, vol. 17, Stanford, CA, USA, 2000, pp. 591–598.

[16] C. Pinho, A. Oliveira, C. Jácome, J. Rodrigues, and A. Marques, "Automatic crackle detection algorithm based on fractal dimension and box filtering," *Procedia Comput. Sci.*, vol. 64, pp. 705–712, Oct. 2015.

[17] S. Emrani, T. Gentimis, and H. Krim, "Persistent homology of delay embeddings and its application to wheeze detection," *IEEE Signal Process. Lett.*, vol. 21, no. 4, pp. 459–463, Apr. 2014.

[18] I. Sen, M. Saraclar, and Y. P. Kahya, "A comparison of SVM and GMM-based classifier configurations for diagnostic classification of pulmonary sounds," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 7, pp. 1768–1776, Jul. 2015.

[19] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, Feb. 2009.
[20] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, 2006.
[21] R. X. A. Pramono, S. Bowyer, and E. Rodriguez-Villegas, "Automatic adventitious respiratory sound analysis: A systematic review," *PLoS ONE*, vol. 12, no. 5, May 2017, Art. no. e0177926.
[22] M. Aykanat, O. Kilic, B. Kurt, and S. Saryal, "Classification of lung sounds using convolutional neural networks," *EURASIP J. Image Video Process.*, vol. 1, pp. 65–74, Sep. 2017.
[23] J. Gu *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, May 2018.
[24] D. Bardou, K. Zhang, and S. M. Ahmad, "Lung sounds classification using convolutional neural networks," *Artif. Intell. Med.*, vol. 88, pp. 58–69, Jun. 2018.
[25] R. G. Stockwell, L. Mansinha, and R. P. Lowe, "Localization of the complex spectrum: The S transform," *IEEE Trans. Signal Process.*, vol. 44, no. 4, pp. 998–1001, Apr. 1996.
[26] P. D. McFadden, J. G. Cook, and L. M. Forster, "Decomposition of gear vibration signals by the generalised S transform," *Mech. Syst. Signal Process.*, vol. 13, no. 5, pp. 691–707, Sep. 1999.
[27] A. Moukadem, A. Dieterlen, and C. Brandt, "Shannon entropy based on the S-transform spectrogram applied on the classification of heart sounds," in *Proc. ICASSP*, New York, NY, USA, May 2013, pp. 704–708.
[28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[29] C. Wachinger, M. Reuter, and T. Klein, "DeepNAT: Deep convolutional neural network for segmenting neuroanatomy," *NeuroImage*, vol. 170, pp. 434–445, Apr. 2018.
[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.
[31] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. ICML*, Haifa, Israel, 2010, pp. 807–814.
[32] B. M. Rocha *et al.*, "A respiratory sound database for the development of automated classification," in *Precision Medicine Powered by pHealth and Connected Health*. Singapore: Springer, 2018, pp. 33–37.
[33] N. Sengupta, M. Sahidullah, and G. Saha, "Lung sound classification using cepstral-based statistical features," *Comput. Biol. Med.*, vol. 75, pp. 118–129, Aug. 2016.
[34] A. Mondal, P. Banerjee, and H. Tang, "A novel feature extraction technique for pulmonary sound analysis based on EMD," *Comput. Methods Programs Biomed.*, vol. 159, pp. 199–209, Jun. 2018.

HAI CHEN (M'18) received the B.S. degree in electrical engineering from the Shaanxi University of Science and Technology, Xi'an, China, in 1995, and the M.S. degree in logistics engineering from Beijing Wuzi University, Beijing, China, in 2008. She is currently pursuing the Ph.D. degree in computer science and engineering with the Macau University of Science and Technology, Macau. Since 2007, she has been an Associate Professor with the School of Information Technology, Beijing Normal University, Zhuhai, China. Her research interests include signal processing, speech recognition, machine learning, deep learning, and the Internet of Things.

XIAOCHEN YUAN (S'08–M'14) received the B.Sc. degree in electronic information technology from the Macau University of Science and Technology, in 2008, and the M.Sc. degree in e-commerce technology and the Ph.D. degree in software engineering from the University of Macau, in 2010 and 2013, respectively, where she was a Postdoctoral Fellow with the Department of Computer and Information Science, from 2014 to 2015. She is currently an Assistant Professor with the Faculty of Information Technology, Macau University of Science and Technology. Her research interests include digital multimedia processing, digital watermarking, multimedia forensics, and deep learning techniques and applications.

ZHIYUAN PEI is currently pursuing the B.S. degree in electronic information technology with Beijing Normal University, Zhuhai, China. His research interests include signal processing, machine learning, and deep learning.

MIANJIE LI is currently pursuing the Ph.D. degree in electronic information technology with the Macau University of Science and Technology, Macau, China. His research interests include digital multimedia processing and digital watermarking.

JIANQING LI received the Ph.D. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 1999. From 2000 to 2002, he was a Visiting Professor with Information and Communications University, Daejeon, South Korea. From 2002 to 2004, he was a Research Fellow with Nanyang Technological University, Singapore. He joined the Macau University of Science and Technology, Macau, in 2004, where he is currently a Professor. His research interests include wireless networks, fiber sensors, and the Internet of Things.