Automatic Classification of Respiratory Sounds Based on Convolutional Recurrent Neural Network and Bagging *k*-Nearest Neighbor

Koki Minami[1], Huimin Lu [1], Tohru Kamiya [1,*], and Shoji Kido [2]

[1] Department of Mechanical and Control Engineering, Faculty of Engineering, Kyushu Institute of Technology, Kitakyushu City, Japan
[2] Osaka University Institute for Radiation Science, Osaka University, Osaka, Japan
Email: riku@cntl.kyutech.ac.jp (K.M.); kamiya@cntl.kyutech.ac.jp (T.K.); kido@radiol.med.osaka-u.ac.jp (S.K.)
*Corresponding author

Abstract—**Respiratory diseases or lung diseases such as asthma bronchiectasis cystic fibrosis are a serious disease. Approximately 8 million people died in each year by chronic obstructive pulmonary disease, lower respiratory tract infections, trachea, bronchial and lung tumors. In addition, COVID-19 is prevalent worldwide in recent years. To analyze these symptom, auscultation of respiratory sounds is very important for screening the respiratory disease. However, there is no quantitative evaluation method for the diagnosis of respiratory sounds until now. To overcome this problem, it is necessary to develop a system to support the diagnosis of respiratory sounds. In the development of support system for auscultation, research by a large-scale, open database used in ICBHI (The International Conference on Biomedical and Health Informatics) 2017 Challenge is in progress. It is expected that a general purpose and highly accurate system will be developed using this dataset. We describe an algorithm for the automatic classification of the respiratory sounds as crackles, wheeze, both, and normal. We improve the classification rates compared with other ICBHI 2017 Challenge teams based on three components. First, we generate the spectrogram images by short-time Fourier transformation. We also extract features using a convolutional recurrent neural network. Third, we classify unknown respiratory sounds by bagging *k*-nearest neighbor algorithm. In the experiment, we applied our proposed method to 920 respiratory sound data which is obtained by the ICBHI Challenge data sets, and achieved *Sensitivity* with 0.670, *Specificity* with 0.863, *ICBHI Score* with 0.766 respectively. Also, area under the curve based on receiver operating characteristic curve of normal class with 0.892, crackle with 0.891, wheeze with 0.874, both with 0.883 were obtained respectively.**

Keywords—**respiratory sounds classification, computer aided diagnosis, short-time Fourier transform, convolutional recurrent neural network, k-nearest neighbor algorithm**

I. INTRODUCTION

Respiratory diseases or lung diseases such as asthma bronchiectasis and cystic fibrosis are serious diseases. It is pathological conditions affecting the organs that make gas exchange difficult in air breathing. According to the survey of the World Health Organization, respiratory disease occupies three out of the top ten causes of death in 2019 (3rd: Chronic obstructive pulmonary disease, 4th: Lower respiratory infections, 6th: Trachea, bronchus, lung cancer), and approximately 8 million died in each year [1]. Furthermore, the COVID-19 is prevalent worldwide in recent year [2]. To reduce the number of deaths due to respiratory diseases, it is important to detect those diseases in early stage and treat them properly.

The auscultation of respiratory sounds with a stethoscope has been widely used as inexpensive, non-invasive and safe, besides taking less time method for the diagnosis of respiratory sounds. However, there is no quantitative evaluation method. Also, auscultation is a subjective process that depends on own hearing of the doctor, experience, and ability to differentiate between different respiratory sound patterns [3]. Therefore, a reliable and quantitative diagnosis support method such as CAD (Computer-Aided Diagnosis) system is necessary. The CAD can digitize from a medical data, quantifies and analyzes medical data by computer, and presents the output result as a “second opinion” [4].

In recent years, there are some reports using the database used in ICBHI 2017 (The International Conference on Biomedical and Health Informatics 2017) Challenge [5, 6] in the development of support systems for auscultation. Since this database is large scale and open data, it is expected that a general-purpose and highly accurate system will be developed using this database. In ICBHI 2017 Challenge, method using MFCC (Mel-Frequency Cepstrum Coefficient) and HMM (Hidden Markov Model) [7], a method using STFT (Short-Time Fourier Transform), wavelet transform and SVM (Support Vector Machine) [8] were proposed, and satisfactory classification rates are achieved until now. Also, a method using a deep learning technique based on CNN (Convolutional Neural Network) using VGG16 with SVM [9] and ResNet [10], outperformed the ICBHI

Manuscript received May 21, 2024; revised June 10, 2024; accepted July 29, 2024; published January 27, 2025.

2017 Challenge. Feature extraction by deep learning is also useful for respiratory sounds analysis. However, further improvement of accuracy is needed for the development of the CAD system.

In this paper, we describe the development of a CAD system for the automatic classification of the large-scale respiratory sound database. Our approach converts respiratory sound data which are one-dimensional signal into two-dimensional images and inputs images which show the characteristics of respiratory sound to CRNN (Convolutional Recurrent Neural Network). Furthermore, the respiratory sounds are classified Bagging *k*-NN (*k*-Nearest Neighbor algorithm) using the features extracted from CRNN.

II. Methods

We classify respiratory sounds by converting the input respiratory sound data into two-dimensional images. For indemnification of respiratory sound, converted images are given to CRNN classifier with Bagging *k*-NN learning. Fig. 1 shows an overview of the CAD system. Details are shown below.

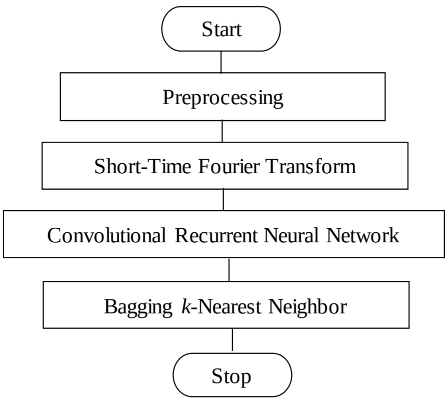


Fig. 1. The flow of CAD system.

A. *Preprocessing*

Respiratory sounds used in ICBHI 2017 Challenge recorded different sampling rates such as 44,100, 10,000, and 4000 Hz. Therefore, each respiratory sound segment was resampled to 4000 Hz. Afterwards, normalized the volume of respiratory sounds.

B. *STFT (Short-Time Fourier Transform)*

Signal conversion was performed on respiratory sound data that has been preprocessed. In this paper, we generated spectrograms to convert respiratory sound to image. Short Time Fourier Transform (STFT) is one of the basic signal processing techniques that has long been used for time-frequency analysis of acoustic signals. It is a method for finding the time change of the frequency and frequency component of local signals.

STFT $F_w(w,\tau)$ is defined by shifting τ the window function $w(t)$ with finite time length L to the original signal $f(t)$ and applying Fourier transform to it.

$$F_w(w,\tau) = \int_{-\infty}^{\infty} w(t-\tau) f(t) e^{-jwt} dt \tag{1}$$

In Eq. (1), t is time. The window function has the role of cutting out the original signal. In this paper, we used the Hamming window as the window function.

The square of the absolute value of $F_w(w,\tau)$ is the power spectrum. Expressing the power spectrum on a two-dimensional plane ($\tau - w$ plane: time-frequency plane), it is possible to visually grasp how the power spectrum changes with time. Such diagrams are called spectrogram [11]. In this paper, we set L to 40 ms and τ to 20 ms. Fig. 2 shows an example of the created spectrogram. Each spectrogram image tends to have a different characteristic on the images.

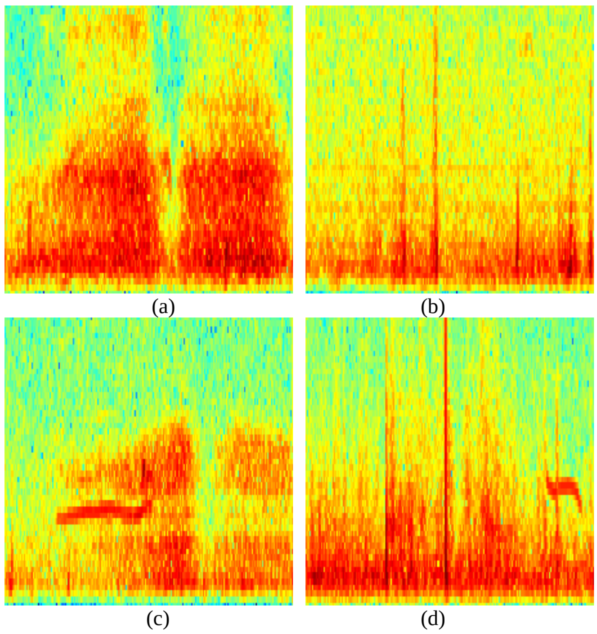


Fig. 2. Example of spectrogram: (a) shows a normal sound, (b) shows crackle sound, (c) shows wheeze sound, and (d) shows both (crackle+wheeze) sound, respectively.

C. *CRNN*

Convolutional Recurrent Neural Network (CRNN) is a model that combines CNN and Recurrent Neural Network (RNN). By incorporating RNN, we expect learning to take time series information into account. In fact, CRNN outperformed CNN in music classification [12].

CNN is a forward propagation network that is mainly used in image recognition. A typical CNN has convolutional layers and pooling layers. The role of convolutional layer is to extract the characteristic shading structures from the image. The convolution operation is shown in Eq. (2).

$$u_i^l = \sum_i z_i^{l-1} h_i^l + b_i^l \tag{2}$$

In Eq. (2), z_i^{l-1} shows previous layer image, h_i^l is filter and b_i^l is bias, respecively.

We apply activation function to u_i^l. Rectified Linear Unit (ReLU) prevents gradient explosion and gradient disappearance problems. The ReLU activation function is shown in Eq. (3).

$$f(u_i^l) = \max(u_i^l, 0) \tag{3}$$

The role of pooling layer is to reduce the amount of information while retaining the important features. Typical pooing includes maximum pooling and average pooling. In this paper, we use maximum pooling which is shown in Eq. (4).

$$u_i^l = \max z_i^{l-1} \tag{4}$$

Fully connected layer tightly bonds all units in adjacent layers. CNN mainly uses ReLU and Softmax function as activation function for fully connected layer. Softmax function is an activation function for classification.

$$y_k^l = \frac{e^{u_k^{(L)}}}{\sum_{j=1}^{K} e^{(u_j^{(L)})}} \tag{5}$$

Batch normalization layer performs normalization for each mini-batch to prevent gradient vanishing during the training stage. Batch normalization can improve learning speed and suppress overfitting [13]. The operation of batch normalization layer is shown in Eq. (6) to (9).

$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{6}$$

$$\sigma_B{}^2 = \frac{1}{m}\sum_{i=1}^{m} (x_i - \sigma_B)^2 \tag{7}$$

$$\hat{x}_l = \frac{x_i - \mu_B}{\sqrt{\mu_B{}^2 + \varepsilon}} \tag{8}$$

$$y_k^l = \gamma \hat{x}_l + \beta \tag{9}$$

In Eq. (6) to (9), $B = \{x_1, x_2, \mathrm{L}\ , x_m\}$ is mini-batch, m is input size, μ_B is average of B, $\sigma_B{}^2$ is variance of B, ε is the small constant, and γ, B is the scale and shift parameter respectively.

RNN is a neural network that handles time series data such as audio and video images. The RNN can capture and estimate the context of time series. However, the length of the time series that can be output to reflect the stored information is short. Long Short-Term Memory (LSTM) solves this problem and makes it possible to handle long-term time series data [14, 15]. The LSTM has three units; input gate, output gate, and forget gate that allow for efficient propagation of errors. The output of each unit is shown in Eq. (10) to (15).

$$z_t = \tanh(W_z x_t + R_z h_{t-1} + b_z) \tag{10}$$

$$i_t = \sigma(W_i x_t + R_i h_{t-1} + p_i \otimes c_{t-1} + b_i) \tag{11}$$

$$f_t = \sigma(W_f x_t + R_f h_{t-1} + p_f \otimes c_{t-1} + b_f) \tag{12}$$

$$c_t = i_t \otimes z_t + c_{t-1} \otimes f_t \tag{13}$$

$$o_t = \sigma(W_o x_t + R_o h_{t-1} + p_o \otimes c_t + b_o) \tag{14}$$

$$h_t = \tanh(c_t) \otimes o_t \tag{15}$$

In Eq. (10) to (15), W,R,b,p is weights, x_t is input, h_t is output, and σ is sigmoid function, respectively.

In supervised learning, it is done by minimizing the classification error of the labeled data. In multiple class classification, the classification error is obtained by the cross-entropy function which is shown in Eq. (16).

$$C = -\sum_{k=1}^{n} d_k \log y_k \tag{16}$$

here, y_k is output of softmax function, and d_k is ideal output (if k is correct class, d_k=1, otherwise 0). We learn to minimize C. For initialization, a random value generated from a normal distribution with mean 0 and variance $\sqrt{2/n}$ for number of nodes n is used as the value [16], and the bias is set to 0.

Stochastic Gradient Descent (SGD) is used to optimize the parameters and minimize the error function. The SGD improves computational efficiency and reduces the possibility of falling into a local solution. Eq. (17) shows calculating the gradient of the error function ∇E_n and updating the weights w.

$$w^{(t+1)} = w^{(t)} - \varepsilon \nabla E_n \tag{17}$$

here, ε is learning rate. In general, the SGD performs weight updates for each mini-batch to improve computational efficiency which is shown in Eq. (18).

$$E_t(w) = \frac{1}{N_t}\sum_{n \in B_t} E_n(w) \tag{18}$$

here, B_t is mini-batch, and N_t is the number of samples included in B_t. In updating the weights, the error is calculated, and the parameters are updated in the direction of its gradient.

Regularization is a method to mitigate overfitting by constraining the degrees of freedom of the weights during learning. A typical method is L2 regularization which adds the sum of the squares of the weights to the error function. L2 regularization is shown in Eq. (19).

$$E_t(w) = \frac{1}{N_t}\sum_{n \in B_t} E_n(w) + \frac{\lambda}{2}\|w\|^2 \tag{19}$$

where λ is parameter that controls the regularization. L2 regularization is called weight decay.

Dropout is a method that learns by probabilistically selecting units in a multilayer network. It reduces the degree of freedom and mitigates overfitting. During training, the units in each layer are randomly selected in proportion p, and the other units are disabled. During inference all units are used in the forward propagation calculation. However, the output of each neuron is multiplied by the proportion p.

In this paper, we created a feature extractor based on CNN with 5 convolutional layers and 3 pooling layers and change fully connected layer to LSTM. Details are shown below. Table I shows the structure of our CRNN.

TABLE I. DETAIL OF THE PROPOSED NETWORK; BN: BATCH NORMALIZATION, RELU: RECTIFIED LINEAR UNIT, DECAY: L2 REGULARIZATION

Layer	Size/Stride	Output	Activation
Input	-/-	64×64×3	-
Conv. 1	3×3/1	64×64×64	BN, ReLU, Decay = 0.0002
Max Pool. 1	2×2/2	32×32×64	-
Conv. 2	3×3/1	32×32×128	BN, ReLU, Decay = 0.0002
Conv. 3	3×3/1	32×32×128	BN, ReLU, Decay = 0.0002
Max Pool. 2	2×2/2	16×16×128	-
Conv. 4	3×3/1	16×16×256	BN, ReLU, Decay = 0.0002
Conv. 5	3×3/1	16×16×256	BN, ReLU, Decay = 0.0002
Max Pool. 3	2×2/2	8×8×256	-
Reshape	-	8×2048	-
LSTM 1	-	8×1024	BN, Dropout = 0.1
LSTM 2	-	1024	BN, Dropout = 0.1
FC	-	4	Softmax

D. *Bagging k-Nearest Neighbor*

Bagging (Bootstrap AGGregatING) is one of the ensemble methods [17, 18]. The ensemble method combines weak learners to create a strong classifier, and bagging uses the bootstrap method. Aggregate the weak classifier and determine the classification class by majority vote. Given m training datasets, we sample m with replacement for each data with probability $1/m$. By repeating this process T times, we can obtain T samples containing m training data. In this paper, we set T to 50.

III. RESULT AND DISCUSSION

A. *Detail of Dataset*

ICBHI 2017 Challenge Dataset is used to evaluate the proposed method. It has collected 920 audio data from 126 patients. The audio files are divided by respiratory cycle and tagged as Normal, Crackle, Wheeze, and Both. The number of respiratory cycles for each class and the total number of cycles are shown in Table II.

TABLE II. NUMBER OF CYCLE

Class	**Number**
Normal	3642
Crackle	1864
Wheeze	886
Both	506
Total	6896

B. *Evaluation Method and Criteria*

In this paper, we evaluate by 5-fold cross validation. The evaluation indices used are *Sensitivity*, *Specificity*, *ICBHI Score*, and AUC (Area Under the Curve) based on the ROC (Receiver Operation Characteristic) curve for each class [19]. *Sensitivity*, *Specificity*, and *ICBHI Score* are shown in Eqs. (20) to (22). Table III shows the determination rules employed to calculate *Sensitivity* and *Specificity*. In addition, a two-sided t-test at significance level of 0.05 is performed on the AUC of the bagging k-NN model and the other models.

TABLE III. DETERMINATION RULES;(N: NORMAL, C: CRACKLE, WHEEZE: WHEEZE, B: BOTH)

Reference label	**Entry's output**			
	C	***W***	***B***	***N***
C	C_c	C_w	C_b	C_n
W	W_c	W_w	W_b	W_n
B	B_c	B_w	B_b	B_n
N	N_c	N_w	N_b	N_n

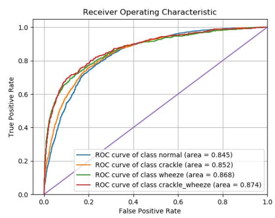
$$Sensitivity = \frac{C_c + W_w + B_b}{C + W + B} \tag{20}$$

$$Specificity = \frac{N_n}{N} \tag{21}$$

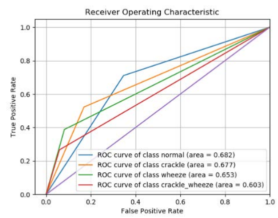
$$ICBHI\ Score = \frac{Sensitivity + Specificity}{2} \tag{22}$$

C. *Experimental Results*

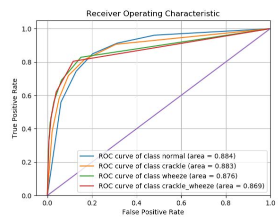
Table IV shows the results of *Sensitivity*, *Specificity*, and *ICBHI Score*. Table V shows the AUC score, and Table IV shows the p-value of the two-sided t-test. Bagging k-NN shows *Sensitivity* of 0.670, *Specificity* of 0.863, and *ICBHI Score* of 0.766. It also shows AUC score of 0.892 for Normal class, 0.891 for Crackle class, 0.889 for Wheeze class and 0.883 for Both classes. Also, satisfactory experimental results are obtained in p-value of the two-sided t-test. Fig. 3 shows ROC which is obtained by each classifier. Among them, proposed Bagging k-NN achieved best performance on normal, crackle, wheeze, and both.



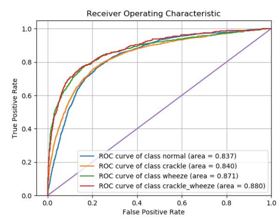
(a) Softmax



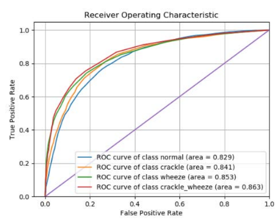
(b) Decision tree



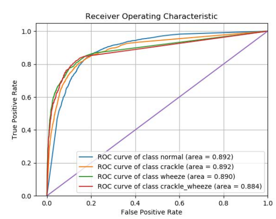
(c) K-NN



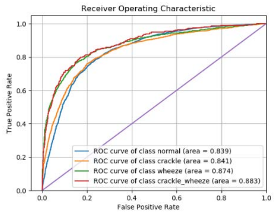
(d) SVM



(e) Random forest



(f) Bagging k-NN



(g) Bagging SVM

Fig. 3. ROC curve.

TABLE IV. RESULTS OF *SENSITIVITY*, *SPECIFICITY* AND *ICBHI SCORE*

Model	*Sensitivity*	*Specificity*	*ICBHI Score*
Softmax (only CRNN)	0.637	0.750	0.694
Decision tree	0.446	0.710	0.578
k-NN	0.651	0.868	0.759
SVM	0.593	0.808	0.700
Random forest	0.469	**0.870**	0.670
Bagging *k*-NN (Proposed method)	**0.670**	0.863	**0.766**
Bagging SVM	0.590	0.808	0.699
CNN+LDA-RSE [20]	0.576	0.832	0.704
RNN [21]	0.584	0.730	0.657

TABLE V. RESULT OF AUC SCORE

Model	Normal	Crackle	Wheeze	Both
Softmax (only CRNN)	0.844	0.851	0.868	0.874
Decision tree	0.682	0.676	0.653	0.603
k-NN	0.883	0.882	0.876	0.869
SVM	0.837	0.839	0.871	0.880
Random forest	0.828	0.841	0.853	0.863
Bagging *k*-NN(Proposed method)	**0.892**	**0.891**	**0.889**	**0.883**
Bagging SVM	0.839	0.840	0.874	**0.883**

TABLE VI. P-VALUE OF THE TWO-SIDED T-TEST

Model	Normal	Crackle	Wheeze	Both
Softmax (only CRNN)	3.69×10^{-5}	3.26×10^{-6}	0.0594	0.260
Decision tree	6.14×10^{-6}	1.92×10^{-5}	1.91×10^{-5}	8.18×10^{-6}
k-NN	5.10×10^{-3}	0.0181	0.0102	5.53×10^{-3}
SVM	1.03×10^{-5}	1.37×10^{-6}	0.0513	0.303
Random forest	5.45×10^{-6}	2.45×10^{-5}	3.75×10^{-4}	5.24×10^{-3}
Bagging SVM	9.27×10^{-6}	1.83×10^{-6}	0.0705	0.790

D. Discussions

k-NN and Bagging *k*-NN classifier achieved the best performance on classification. In the feature map obtained from CRNN, similar features are mixed, but we believe that *k*-NN, which decides classification by majority vote, was not affected by the similar features. The two-dimensional embedding space just before the final output layer of CRNN is shown in Fig. 4.

Bagging *k*-NN shows better AUC than *k*-NN and below the significance level of 0.05, so using bagging is significant. Further improvement in accuracy can be expected by adjusting the hyperparameters and number of weak learners. On the other hand, there are models in which significant differences cannot be confirmed in Wheeze and Both classes. Wheeze and Both classes have less data than the other classes. We believe that a better model can be created by applying the data expansion method to audio data or spectrogram images.

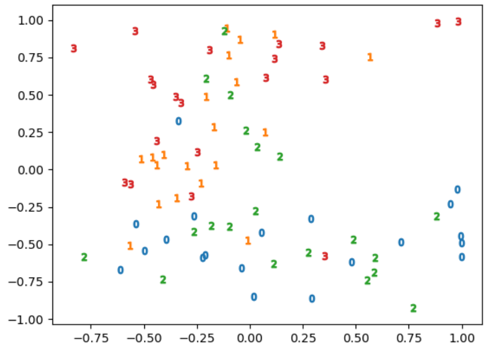


Fig. 4. Embedding space;(0: Normal, 1: Crackle, 2: Wheeze, 3: Both).

IV. CONCLUSION

In this paper, we proposed a CAD system to classify respiratory sounds from large-scale respiratory sound database. Spectrogram images were generated by STFT on ICBHI 2017 Challenge database. After training CRNN, we used the features obtained from CRNN to classify the respiratory sounds from the unknown data using Bagging *k*-NN. The performance of classifier was *Sensitivity* of 0.670, *Specificity* of 0.863, and *ICBHI Score* of 0.766, and AUC of 0.892 for Normal class, 0.891 for Crackle class, 0.889 for Wheeze class and 0.883 for Both classes. Demir *et al.* [20] propose a classification of lung sounds based on CNN model using parallel pooling structure and obtained *Sensitivity* of 0.61, *Specificity* of 0.86, and *ICBHI Score* of 0.735. Our new method using the ICBHI Score was improved by 3.1%, when compared to the other best result methods using the same dataset [21]. In our study, we have limitation that "Wheeze" and "Both" categories are relatively small. To overcome this problem, GAN (Generative Adversarial Network)-based data generation methods should be considered. Although some principles for selecting network structures and hyper-parameters are described in this article, the actual model performance may be affected by the selection of hyper-parameters. Methods for selecting the optimum hyper-parameters need to be considered, and these are also future tasks. In this study, we evaluated the accuracy against the ICBHI open dataset. We plan to conduct classification experiments using clinical data as well to evaluate its performance in the future.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Koki Minami conducted the research and wrote the initial manuscript, and Tohru Kamiya supervised the project, provided suggestions and recommendations along the way. Huimin Lu and Shoji Kido give revised manuscripts. All authors had approved the final version.

ACKNOWLEDGEMENTS

In this paper, we used ICBHI 2017 Challenge Dataset (https://bhichallenge.med.auth.gr/ICBHI_2017_Challenge).

REFERENCES

[1] World Health Organization. (December 2020). The top 10 causes of death. [Online]. Available: https://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death

[2] N. Zhu *et al.*, "A novel coronavirus from patients with pneumonia in China," *The New England Jornal of Medicine*, vol. 382, no. 8, pp. 727–733, 2020.

[3] A. Kandaswamy *et al.*, "Neural classification of lung sounds using wavelet coefficients," *Computers in Biology and Medicine*, vol. 34, no. 6, pp. 523–537, 2004.

[4] K. Doi, "Computer-aided diagnosis in medical images: Historical review, current status and future potential," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4–5, pp. 198–211, 2007.

[5] ICBHI 2017 Challenge. (January 2021). Respiratory sound database. [Online]. Available: https://bhichallenge.med.auth.gr/ICBHI_2017_Challenge

[6] B. M. Rocha *et al.*, "An open access database for the evaluation of respiratory sound classification algorithms," *Physiological Measurement*, vol. 40, no. 3, 035001, 2019.

[7] N. Jakovljević and T. Lončar-Turukalo, "Hidden markov model based respiratory sound classification algorithms," P*recision Medicine Powered by pHealth and Connected Health*, pp. 39–43, 2018.

[8] G. Serbes *et al.*, "An automated lung sound preprocessing and classification system based on spectral analysis methods," *International Conference on Biomedical and Health Informatics*, pp. 45–49, 2018.

[9] F. Demir *et al.*, "Convolutional neural networks based on efficient approach for classification of lung diseases," *Health Information Science and Systems*, vol. 8, no. 1, pp. 1–8, 2020.

[10] H. Chen *et al.*, "Triple-classification of respiratory sounds using optimized s-transform and deep residual networks," *IEEE Access*, vol. 7, pp. 32845–32852, 2019.

[11] J. Allen, "Short term spectral analysis, synthesis, and modification by discrete fourier transform," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 25, no. 3, pp. 235–238, 1977.

[12] K. Choi *et al.*, "Convolutional recurrent neural network for music classification," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2017, pp. 2392–2396.

[13] S. Loffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," arXiv Print, arXiv: 1502.03167, 2015.

[14] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[15] K. Greff *et al.*, "LSTM: A search space odessey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, 2017.

[16] K. He *et al.*, "Delving deep int rectifiers: Surpassing human-level performance on ImageNet classification," *IEEE International Conference on Computer Vision*, pp. 1026–1034, 2015.

[17] B. V. Dasarathy, "Nearest Neighbor (NN) norms: NN pattern classification techniques," *IEEE Computer Society Tutorial*, 1991.

[18] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, pp. 123–140, 1996.

[19] A. K. Akobeng, "Understanding diagnostic Tests 3: Receiver operating charcteristic curves," vol. 96, no. 5, pp. 644–647, 2007.

[20] F. Demir *et al.*, "Classification of lung sounds with CNN model using parallel pooling structure," *IEEE Access*, vol. 8, pp. 105376–105383, 2020.

[21] K. Kochetov *et al.*, "Noise masking recurrent neural network for respiratory sound classification," *International Conference on Artificial Neural Networks*, pp. 208–217, 2018.