

Automatic Classification of Respiratory Sounds by Improving the Loss Function of ResNet

Ryusei Oshima

Kyushu Institute of Technology, 1-1 Sensui, Tobata, Kitakyushu, 804-0015, Japan

Tohru Kamiya

Kyushu Institute of Technology, 1-1 Sensui, Tobata, Kitakyushu, 804-0015, Japan

Shoji Kido

Osaka University, 2-2 Yamadaoka, Suita, Osaka, 565-0871, Japan
Email: oshima.ryusei710@mail.kyutech.jp, kamiya@cntl.kyutech.ac.jp

Abstract

Respiratory diseases cause 8 million deaths annually, and this number is expected to increase. Breath auscultation, a primary diagnostic method, is noninvasive, repeatable, and immediate, but faces challenges such as reliance on skilled practitioners, difficulty in quantitative assessment, and limited accessibility in developing regions or disaster sites. To address these issues, we developed a deep learning-based breath sound classification system using the ICBHI 2017 dataset. Our method classifies breath sounds into four categories: Normal, Crackle, Wheeze, and Crackle and Wheeze. We use ResNet-34 as the base model, which is enhanced with CBAM for better spatial and channel feature extraction. To deal with class imbalances, we incorporate Focal Loss. The system achieves *Accuracy* of 0.732, *SE* of 0.607, *SP* of 0.843, and *ICBHI Score* of 0.725.

Keywords: Respiratory Sounds, Convolutional Neural Network, ResNet, CBAM, Focal Loss

1. Introduction

Respiratory diseases include many types of diseases such as tumors, infectious diseases, allergies, and autoimmune diseases, and bronchial asthma, COPD (Chronic Obstructive Pulmonary Disease), lung cancer, and respiratory tract infections (bronchitis, pneumonia, etc.) are considered the major respiratory diseases [1]. Pneumonia, in particular, is the third leading cause of death among Japanese people. The main cause of pneumonia in the elderly is infection with *Streptococcus pneumoniae*, a type of bacteria that normally lives in the mouth and on the skin, and which rarely causes infection in healthy people [2].

In addition, chronic obstructive pulmonary disease is the third leading cause of death worldwide in 2019, lower respiratory tract infections are the fourth leading cause, and cancers of the trachea, bronchus, and lungs are the sixth leading cause, with about 8 million deaths due to respiratory diseases each year [3]. The increase of respiratory diseases is remarkable worldwide, and WHO (World Health Organization) predicts that COPD, respiratory tract infections, and respiratory tract cancer will be the third to fifth leading causes of death in the world. COPD, respiratory infections, and respiratory cancers are expected to account for the fifth to sixth largest number of deaths. Therefore, early detection and treatment are expected to reduce the number of deaths from these diseases.

Pulmonary auscultation is the main diagnostic method to identify respiratory diseases. Auscultation is a method of classifying abnormal breath sounds caused by lung and

bronchial diseases by listening to the breath sounds with a stethoscope. The advantages of auscultation are that it is noninvasive, can be performed repeatedly, and the results can be obtained immediately. However, auscultation also has disadvantages: accurate diagnosis requires skill, quantitative evaluation is difficult, and diagnosis is difficult in developing countries where there are not enough doctors or at disaster sites. Therefore, there is a need to develop applications that can quantitatively evaluate and diagnose breath sounds.

The ICBHI 2017 Challenge Dataset is now available. This dataset consists of four classes of breath sounds: Normal, Crackle, Wheeze, and Crackle and Wheeze. Breath sound classification methods have been proposed around the world using this dataset. It contains breath sounds recorded using multiple recording devices, and it is possible to conduct research that takes into account differences in breath sounds due to differences in microphones.

Many breath sound classification methods using deep learning have been proposed in related research. Among them, methods using multi-layered CNN (Convolutional Neural Network) models such as ResNet (Residual Network) have attracted attention in breath sound classification [4]. In this paper, we attempt to construct a deep-learning model that can automatically classify raw breath sounds by extracting their features. The method is based on ResNet [5], and an improved deep-learning model is used for automatic classification. We apply the proposed method to breathe sound data, evaluate its performance in automatic classification, and discuss the results.

2. Methodology

2.1. Preprocessing

The breath sounds in the ICBHI 2017 Challenge Dataset were recorded with different sampling rates (44100 Hz, 10000 Hz, and 4000 Hz) and with different microphones. First, all the breath sound data were resampled to 4000 Hz to align the sampling rates. Second, volume normalization is performed to reduce the effect of volume differences during recording. Third, the breath sound data are clipped at each respiratory cycle and labeled (Normal, Crackle, Wheeze, Both). Finally, we generate a spectrogram by using STFT (short-time Fourier transform). In this paper, the time length N was 256 (64ms), and the frameshift S is assumed to be none.

2.2. ResNet

ResNet (Residual Network) is a model developed by MSRA (Microsoft Research Asia), which won the 2015 ILSVRC (ImageNet Large Scale Visual Recognition Challenge) competition for image recognition accuracy.

ResNet solves the gradient vanishing problem by introducing a Residual Block, which consists of two convolutional layers, a shortcut connection, and an addition operator. We show the architecture of Residual Block (Fig.1). The ResNet is composed of multiple blocks connected in series. In this paper, we modify the Residual Block to increase accuracy.

2.3. CBAM

CBAM (Convolutional Block Attention Module) (Fig.2) is a channel attention mechanism proposed in 2018 for use in deep learning [6]. CBAM estimates separate attention maps along two dimensions (channel and spatial) from the feature maps and multiplies the feature maps with the attention maps to create new feature maps. In other words, it aims to improve the representational capability of the network by focusing on both channel-wise and spatial information. The Spatial Attention Module and Channel Attention Module used in CBAM are described below.

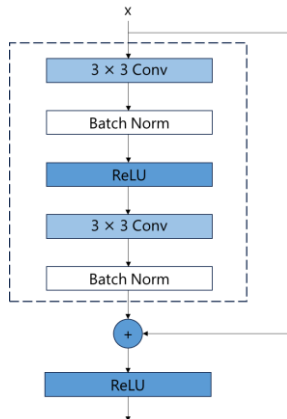


Fig. 1. Residual Block

(i). Channel Attention Module

The Channel Attention Module focuses on channel-wise information and computes the importance scores for each channel.

(ii). Spatial Attention Module

The Spatial Attention Module evaluates the relative importance of different locations within the feature maps and adjusts the feature maps accordingly to emphasize beneficial features.

In this way, CBAM applies channel attention to the feature maps followed by spatial attention, focusing on both channel-wise and spatial information rather than just one, enabling it to extract more important information. Therefore, in this paper, CBAM is incorporated into the ResNet (Fig.3). Specifically, it is introduced after the convolution in the residual blocks. We show CBAM and Residual Block incorporating CBAM in Fig.2 and Fig.3, respectively.

2.4. Focal Loss

Focal Loss is a loss function that acts on class imbalance, aiming to improve model accuracy by focusing on difficult samples [7]. Compared to the commonly used cross-entropy loss, Focal Loss assigns lower weights to easily classified samples and higher weights to difficult samples. This enables the model to focus on important samples, as majority classes get lower weights while minority classes get higher weights. The formula for Focal Loss is shown below.

$$FL(p_t) = -\alpha(1 - p_t)^\gamma \log p_t \quad (1)$$

p_t is the model's estimated probability for the class with a label $0 \leq p_t \leq 1$, and γ is an adjustable parameter. As p_t approaches 1, the loss becomes smaller, exhibiting the characteristic of reinforcing the focus on difficult

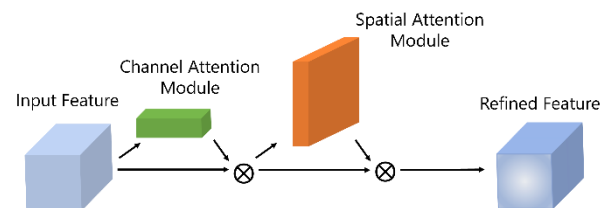


Fig. 2. CBAM

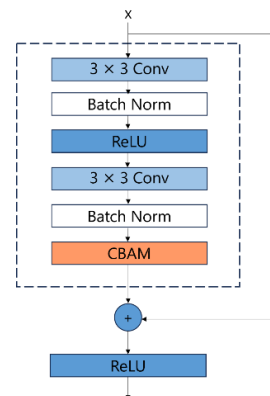


Fig. 3. CBAM is incorporated into the ResNet

samples when p_t is small. When p_t is large, p_t^γ becomes large for samples where p_t is high, meaning the classification is easy, thus suppressing rapid changes. This reduces the loss for majority classes or easily classified samples, making the model more robust to them. The value of γ depends on the task and dataset, so the optimal value needs to be found through experimentation.

In this paper, while CBAM is incorporated into ResNet, the dataset has imbalanced classes with the Normal class constituting half the samples. At this stage, introducing CBAM alone may emphasize the numerous Normal classes. Therefore, introducing Focal Loss enables focusing on the imbalanced classes while mitigating the influence of easy samples.

2.5. Classification

We use ResNet, which has achieved high performance in the field of image classification as a model to extract image features. In this paper, we propose an improved model from the ResNet34.

2.6. Detail of dataset

The experiments in this paper utilized the dataset employed in the ICBHI 2017 Challenge. The dataset used includes four types of respiratory data: Crackle, Wheeze, Both, and Normal, consisting of 920 audio files recorded with an electronic stethoscope from 126 patients (Table 1) (Table 2).

2.7. Result

Table 3 shows the results of the performance evaluation of the proposed method. As a comparison, the results are also shown for the case where only CBAM is added to the Residual Block and where only the loss function is replaced by Focal Loss.

2.8. Discussion

In this paper, ResNet is adopted as the base model, and the accuracy of automatic breath sound classification is improved by modifying ResNet (Table 3). First, we compare the performance of ResNet concerning the number of layers. The experimental results show that ResNet-34 achieves the highest scores in *Accuracy*, *SE*, and *ICBHI Score*. ResNet-34 has more layers than ResNet-18 and can learn more deeply. The results suggest that the deeper training than ResNet-18 enables the extraction of features that are useful for classification. On the other hand, ResNet-18 has a higher score than ResNet-34 in *SP*, but we believe that ResNet-34 is superior to ResNet-18 because a high *SE* is more important for pathological diagnosis, i.e., not to miss a possible disease.

Table 1. Label tree manually constructed for the ICBHI 2017 Challenge dataset

Class	Number
Crackle	1864
Wheeze	886
Both	506
Normal	3642

Table 2. Confusion Matrix of 4 Class

		Prediction Label				Total
		Crackle	Wheeze	Both	Normal	
True Label	Crackle	C_c	C_w	C_b	C_n	C
	Wheeze	W_c	W_w	W_b	W_n	W
	Both	B_c	B_w	B_b	B_n	B
	Normal	N_c	N_w	N_b	N_n	N

Next, we compare the results of introducing CBAM to the Residual Block. CBAM increases the *Accuracy*, *SE*, *SP* and *ICBHI Score* of ResNet-18 and ResNet-34, except for the *SP* of ResNet-18. The introduction of CBAM makes it possible to focus on features in the channel and spatial directions and extract important information in these directions.

Furthermore, we discuss the results of changing the loss function to Focal Loss, which increases the *Accuracy* and *SP* scores for both ResNet-18 and ResNet-34, but decreased the accuracy of the other classes as the accuracy of Normal increases. This may be due to the fact that the attention to the Normal data, which has a relatively large number of data, is distributed to the other classes, thereby reducing overlearning on the Normal data and improving accuracy. As a result, *SP* is improved and *SE* is decreased.

Finally, we compare the experimental results of the model proposed in this paper, in which CBAM and Focal Loss are introduced to ResNet. First, the introduction of CBAM increases the expressive power and enables more precise extraction of important features in each class, but it also decreases the scores for other classes because it focuses too much on features in the Normal data with many images. However, the introduction of Focal Loss and CBAM enables us to shift attention from Normal to other classes, and we believe that this improves overall accuracy.

3. Conclusion

In this paper, we proposed a deep learning model that can classify images transformed by the short-time Fourier transform, using ResNet as the base model, with improvements such as feature extraction using CBAM and weighting of imbalance classes by introducing Focal Loss, and perform automatic classification from breath

Table 3. Result of 4-class classification

Model	Accuracy [%]	SE [%]	SP [%]	ICBHI Score [%]
ResNet-18	0.705	0.584	0.813	0.699
ResNet-34	0.716	0.613	0.807	0.711
ResNet-18 + CBAM	0.715	0.605	0.812	0.709
ResNet-34 + CBAM	0.726	0.623	0.818	0.721
ResNet-18 + Focal Loss	0.717	0.584	0.835	0.710
ResNet-34 + Focal Loss	0.717	0.586	0.834	0.710
ResNet-18 + CBAM + Focal Loss	0.725	0.604	0.832	0.718
ResNet-34 + CBAM + Focal Loss	0.732	0.607	0.843	0.725

sound data in the ICBHI 2017 Challenge Dataset. The results show *Accuracy* of 73.2%, *SE* of 60.7%, *SP* of 84.3%, and *ICBHI Score* of 72.5%, which are better than the base model, ResNet, and also the classification accuracy is improved for classes with small number of data except Normal. In the future, we are considering the use of transfer learning to further improve classification performance. It may be difficult to learn enough features for the classification because the number of data used in this study is too small to fully exploit the power of the deep learning model. Therefore, we believe that transfer learning can be introduced to compensate for the lack of data by appropriating knowledge from models previously trained on other larger data sets. In addition, by using weights from previously trained models as initial values, it is expected that model convergence will be faster and training time will be reduced.

Acknowledgements

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

References

1. World Health Organization, Chronic respiratory diseases, https://www.who.int/health-topics/chronic-respiratory-diseases#tab=tab_1 (2024/06/21 accessed).
2. N. Miyashita, Y. Yamaguchi, Bacterial Pneumonia in Elderly Japanese Populations, 2018, 2018 Jan Japanese Clinical Medicine.
3. World Health Organization, The top 10 causes of death, <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death> (2024/01/15 accessed).
4. S. Gairola, F. Tom, N. Kwatra, M. Jain, RespireNet: A Deep Neural Network For Accurately Detecting Abnormal Lung Sounds In Limited Data Setting, 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2021,
5. K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp.770-778.
6. S. Woo, J. Park, J. Lee, I. Kweon, CBAM: Convolution Block Attention Module, Computer Vision and Pattern Recognition, arXiv:1807.06521[cs.CV], 2018.

7. T. Lin, P. Goyal, R. Girshick, K. He, P. Dollar, Focal Loss for Dense Object Detection, Computer Vision and Pattern Recognition, arXiv:1708.02002[cs.CV], 2018.

Authors Introduction

Mr. Ryusei Oshima



He received his Bachelor's degree in Engineering in 2021 from the Faculty of Engineering, Kyushu Institute of Technology in Japan. He is currently a master student in Kyushu Institute of Technology, Japan.

Dr. Tohru Kamiya



He received his B.A. degree in Electrical Engineering from Kyushu Institute of Technology in 1994, the Masters and Ph.D. degree from Kyushu Institute of Technology in 1996 and 2001, respectively. He is a professor in the Department of Mechanical and Control Engineering at Kyushu Institute of Technology. His research interests are focused on image processing and medical application of image analysis. He is currently working on computer aided diagnosis based on CT, MR imaging, fluorescence microscope imaging, and automatic classification of respiratory sound.

M.D. Ph.D. Shoji Kido



He received his M.D. degree from Osaka University in 1988. He received his Ph.D. degrees in Medicine and Information Science from Osaka University in 1992 and 1999, respectively. He is a guest professor of Osaka University Institute for Radiation Science and Osaka University Graduate School of Medicine. His research interests are focused on the use of artificial intelligence in radiology.
