Application and Performance Improvement of Transfer Learning on ICBHI Lung Sound Dataset

Mohan Xu[1] and Lena Wiese[1,2]

[1] Fraunhofer Institute for Toxicology and Experimental Medicine, Hannover, Germany,
mohan.xu@item.fraunhofer.de,lena.wiese@item.fraunhofer.de
[2] Institute of Computer Science, Goethe University Frankfurt, Frankfurt a. M., Germany

Abstract. Chronic respiratory diseases are one of the leading causes of morbidity and mortality worldwide. How to prevent the disease or to diagnose and treat it effectively in the early stage has always been a focused medical research area. In this paper, a neural network that was pre-trained based on a large audio event dataset called AudioSet is transferred and applied in the training and testing of the Respiratory Sound database ICBHI; in addition, various methods are used in data preprocessing, neural network configuration and post-processing to improve the performance of the transfer learning model. The final model can not only converge quickly, but also use the accuracy calculation method provided by ICBHI Challenge to reach 81.1% in the four classification tasks containing normal, crackle, wheeze and both respiratory sounds, which is superior to the previous methods. This paper also analyzes the unbalanced distribution of the respiratory cycle dataset based on demographic data on the binary classification task (normal and abnormal). The binary classification model scored 85.5% and 81.1% on the female test group and the male test group, respectively. To address the above differences due to the unbalanced dataset, we used a restricted mixup approach to successfully reduce the difference between the male and female test groups to 0.82%.

Keywords: ICBHI dataset, Respiratory sounds classification, Neural network, Transfer learning

1 Introduction

According to a report released by the World Health Organization (WHO) [1], in 2016, the number of deaths from non-communicable diseases worldwide accounted for 71% of the total number of deaths. Chronic respiratory disease ranks third among the four leading causes of death from non-communicable diseases. More than 40% of countries have fewer than 10 doctors per 10,000 people. How to prevent a chronic lung the disease or take effective treatment in the early stage of disease is an important research topic in modern medicine. Developing and verifying novel digital health approaches to support the effective detection and diagnosis of lung diseases are hence an important step in order to improve the outcomes for affected patients.

Stethoscopes are widely used worldwide as a non-invasive method for the analysis of lung sounds. They enable medical staff to diagnose possible diseases by auscultation of the lungs (and potentially other organs as for example the heart). The use of stethoscope for medical diagnosis is not only suitable for a wide range of people, but also can quickly obtain test results, thereby winning precious time for patient treatment. However, the use of a stethoscope is very dependent on the doctor's individual hearing abilities and clinical experience. If the diagnosis results are not accurate enough, it may lead to incalculable consequences. Therefore, it is our research purpose to apply machine learning methods on the existing respiratory system sound database to obtain more accurate results in the future diagnosis of respiratory diseases. This cannot only provide general practitioners with machine-aided analysis of diagnostic data from experienced doctors, but may also provide a valuable digital health tool for patients in remote areas or in home environments due to the feasibility of conducting a diagnosis remotely.

1.1 Our Contribution

This work uses a transfer learning approach to apply Wavegram Logmel CNN [13] trained based on Audioset [14] to the ICBHI dataset [2] and save snapshots [15] of the model at different stages as the learning rate changes. In the preprocessing phase of the data, we provide several data augmentation methods: splitting and padding, *nlpaug* library [11], *rollAudio* and *mixup* [12] and compare their effects on the model performance to select the best model configuration. In a 4-classification task containing respiratory sounds of normal, crackle, wheeze and both, the prediction results of the ensemble model obtained by our transfer learning system after 10-fold cross validation achieved a score of 81.1%, outperforming previous methods. Based on the demographic information provided by ICBHI [2], we investigated the distribution of the dataset in terms of gender, age and BMI. The model scores of the female test group and the male test group based on normal/abnormal respiratory cycle were 85.5% and 81.1%, respectively. To reduce discrimination in the male test group due to the uneven distribution of the dataset, we used a restricted mixup approach to reduce the difference in model scores between the male and female test groups to 0.82%.

1.2 Outline of this paper

The paper is outlined as follows. Section 2 describes research work on the classification task of the ICBHI respiratory sound database. The proposed methodology, introduction of ICBHI dataset, official evaluation methods, preprocessing, transfer learning and ensembling steps are depicted in Section 3. Section 4 presents the experiments and results on the proposed transfer learning system. We conclude this work and identify future directions in Section 5.

2 Related Work

In recent years, machine learning has been widely used in the classification of the respiratory system sound database ICBHI [2]. The sound data enters the classifier after different preprocessing methods, and the prediction of the category to which the data belongs is realized through the learning of parameters. The boosted tree model proposed by [3] takes all the features as input and performs multiple iterations to achieve the prediction of breathing cycle. The LungBRN model [4] simultaneously receives the features obtained by the two preprocessing methods, trains them in the Resnet network respectively, and finally multiplies them in the fully connected layer. [5] combines transfer learning (VGG16 pretrained model) and SVM algorithm. In addition, Recurrent Neural Networks (RNNs) have also been used in some works [6–8] for respiratory cycle classification problems.

In contrast to research works [4, 5, 10] that used neural networks pre-trained on the large-scale visual database Imagenet [9], the proposed transfer learning system applies the neural network pre-trained on the large-scale audio dataset google audioset [14] containing 527 sound categories [13], thus enabling the downstream task to learn audio features faster. In the preprocessing stage, preprocessing methods that have performed well in other tasks such as [6, 17, 20] are used in our respiratory sound classification task. Table 3 shows the comparison between our work and some research works with the same dataset division ratio. In addition, this work investigates the distribution of the dataset on gender, age and BMI, and tests the performance differences of the neural network model on uneven datasets. As an example of uneven datasets due to the gender of the subjects, we propose a restricted mixup approach to reduce the resulting discrimination of the model against the male test group. To the best of our knowledge, this work is the first research on the respiratory sound database ICBHI to analyse network model performance based on demographic information.

3 Methodology

This section begins with a detailed description of the ICBHI dataset and the corresponding evaluation criteria. The general approach of our system based on google audioset's pre-trained model Wavegram Logmel CNN [13, 14] to achieve classification prediction of respiratory cycles on normal, crackle, wheeze and both. The proposed system is shown in Figure 1, including different preprocessing methods for the respiratory cycles, transfer learning of Wavegram Logmel CNN [13] and taking model snapshots at different local minima. The details of the transfer learning system are described in the following subsections.

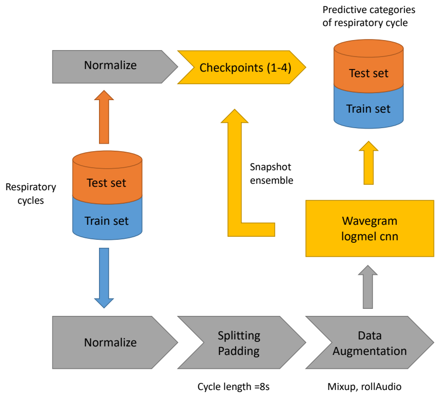


Fig. 1. Workflow of training and testing process on the ICBHI dataset

3.1 Dataset

The ICBHI Scientific Challenge database [2] is a publicly available respiratory sound database that is tested through scientific challenges, thus enabling digital auscultation based on its usability in terms of data and algorithms. This respiratory sound database consists of 920 annotated audio samples from 126 subjects with data categories labeled by respiratory experts as: normal, wheeze, crackle, and both (wheeze and crackle). Each audio sample can be divided into multiple respiratory cycles based on annotation and may contain multiple categories. The database contains a total of 6898 respiratory cycles, of which 1864 contain crackles, 886 contain wheezes, 506 contain both, and the rest are normal. In addition, a large number of samples in the database contain noise, which makes the data classification problem closer to a real-life scenario. Figure 2 shows mel-spectrograms of four representative respiratory cycles from the same subject, belonging to four data categories (normal, wheezing, crackles and both).

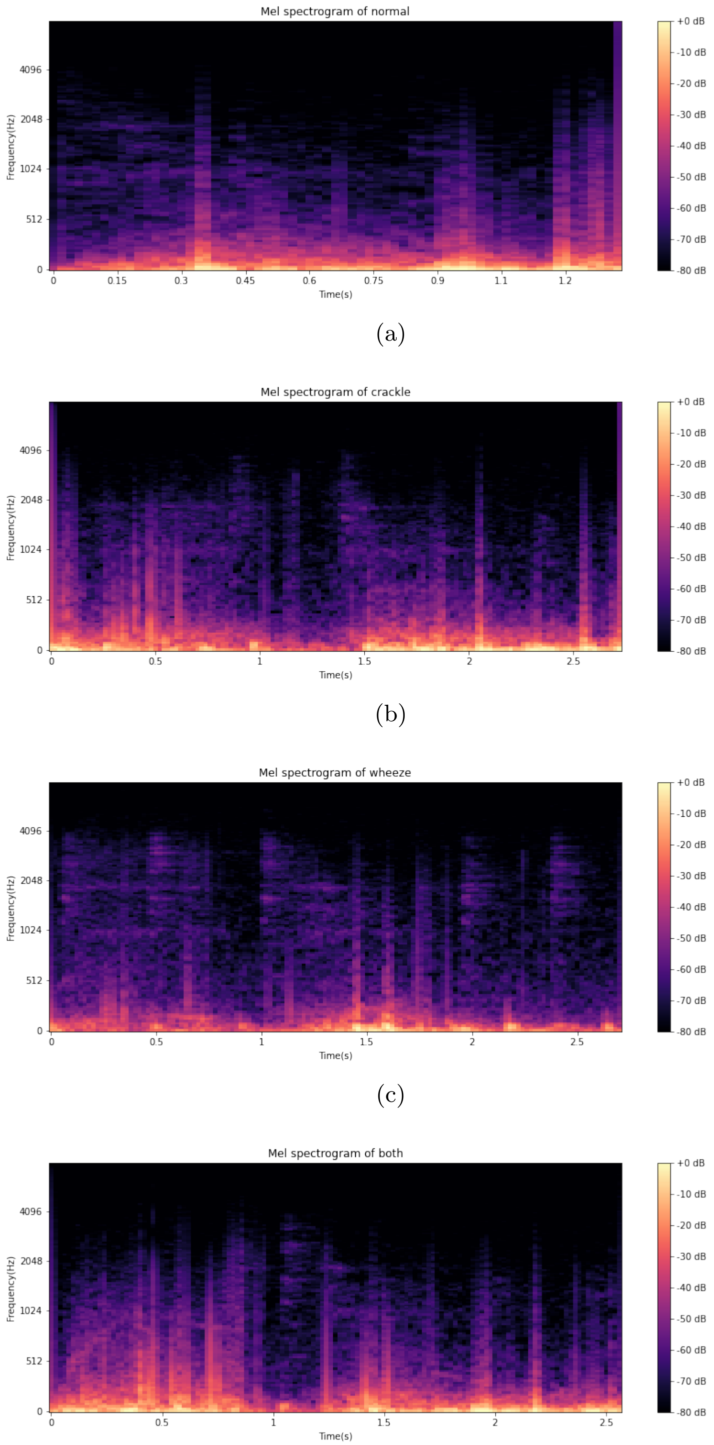
3.2 Evaluation method and criteria

The evaluation method in this work uses the widely used officially proposed criteria. For the four classification (normal (N), crackle (C), wheeze (W) and both (B)) problems, the three measures Sensitivity (S_e), Specificity (S_p) and Score (S_c) are defined as follows,

$$Se = \frac{C_{correct} + W_{correct} + B_{correct}}{C_{total} + W_{total} + B_{total}} \tag{1}$$

$$Sp = \frac{N_{correct}}{N_{total}} \tag{2}$$

$$Sc = (\frac{Se + Sp}{2}) * 100 \tag{3}$$



(a)

(b)

(c)

(d)

Fig. 2. Mel-spectrograms of different classes of respiratory cycles. (a) Normal (b) Crackle (c) Wheeze (d) Both

where *i_correct* and *i_total* denote the number of correctly classified respiratory cycles and the total number of respiratory cycles contained in the class when $i \in \{N, C, B, W\}$.

3.3 Preprocessing

To facilitate the preprocessing of continuous audio signals in the neural network workflow, the librosa library [16] reads sound files and samples them as discrete audio signals at 16 kHz. The sampled data are divided into different respiratory cycles according to the corresponding annotated data and identified to the category (0-3) they belong to. The data for each respiratory cycle are classified according to “data”, “start”, “end”, “label”, “cycle” and “filename” column names (Table 1). The first four columns were used for the whole experiment, i.e., the classification task of the respiratory cycle and the difference in model performance on test groups by gender. The last column was used for the second part of the experiment, where the subject index was identified by the filename corresponding to that respiratory cycle, and thus the subject’s gender information was obtained from the official demographic information provided.

Table 1. Part of the respiratory cycle information contained in a sound file (filename.wav)

Cycle	Start(s)	End(s)	Crackle	Wheeze	Label
1	1.778	4.032	0	0	0
2	4.032	6.319	1	0	1
3	6.319	8.239	0	1	2
4	8.239	10.075	1	1	3

Before starting the formal neural network training, length alignment, data augmentation and normalization operations are required. The following section describes in detail the various data preprocessing methods currently in use.

Splitting and Padding: The respiratory cycle lengths in the ICBHI dataset ranged from 0.2s to 16.1s, while the input shape of the neural network is fixed. Considering the length of the pre-trained dataset in [14, 17] and the performance of the ICBHI dataset trained on different cycle lengths, we set the length of each cycle of the input neural network to 8s.

When the length of a respiratory cycle is greater than 8s, the respiratory cycle will be divided into multiple sub-respiratory cycles, which can be achieved by the framing function of the librosa library [16]. The frame length is set to 8s and the hop length of the frame is 4s. The frame will be displaced in the direction of the respiratory cycle length until the remaining respiratory cycle length does not satisfy the frame length, thus different frames (sub-respiratory cycles) can be obtained.

When the length of a certain respiratory cycle is less than 8s, that respiratory cycle will be repeatedly spliced along the length direction until it is greater than or equal to 8s.

Data Augmentation: In the data preparation phase, a randomly ordered data augmentation combination is formulated, which contains the *NoiseAug*, *SpeedAug*, *LoudnessAug*, *VtlpAug* and *PitchAug* methods from the nlpaug library [11]. The augmented data are processed according to the rollAudio method in Table 2. A random index is generated based on the cycle length, and the data is rolled from this index until it returns to the previous bit of this index.

Table 2. Comparison of arbitrary respiratory cycle before and after rollAudio operation (*3* is the generated random index)

Before	2	3	5	**7**	11	13	17	19
After	**7**	11	13	17	19	2	3	5
Index	0	1	2	**3**	4	5	6	7

mixup [12] augments the dataset by mixing random respiratory cycles over a certain respiratory cycle according to $mixingproportion = (\lambda : 1 - \lambda)$ over the range of the dataset. λ is generated by beta distribution, a set of continuous probability distributions (Equation 4) defined on the (0, 1) interval, by setting alpha and beta and thus controlling the interpolation intensity between two respiratory cycles. For $\alpha, \beta \in (0, \infty)$

$$\lambda \propto Beta(\alpha, \beta), \tag{4}$$

Since this probabilistic event mixup occurs on a pair of respiratory cycles, the interpolated intensity distributions of the two respiratory cycles in mixup are equal in the absence of constraints such as data weights. The beta distribution is uniform only when $\alpha = \beta$ and λ is symmetric on $x = 0.5$. Referring to the experiments of [12], we set both α and β in the beta distribution equal to 0.2 so that the λ used in equations (5) and (6) can be calculated.

$$\tilde{x} = \lambda x_i + (1 - \lambda) x_j \tag{5}$$

$$\tilde{y} = \lambda y_i + (1 - \lambda) y_j \tag{6}$$

(x_i, y_i) are the inputs and targets of the original cycle, respectively, and the inputs and targets of the random cycles involved in the mixup operation are (x_j, y_j), respectively. The results returned by equations (4) and (5) will be used as the inputs of the neural network.

Normalization: If the original data is used directly, the data of different orders of magnitude will have different effects on the analysis results. The data of the larger order of magnitude will weaken the effect of the data of

the smaller order of magnitude in the analysis, which is not the result we want to see. The data processed by Z-score normalization have a mean of 0 and a variance of 1, which are on the same order of magnitude.
The mean and standard deviation corresponding to each respiratory cycle are calculated separately and averaged over the entire dataset. For any data point x we obtain,

$$x' = \frac{x - \mu}{\delta} \tag{7}$$

where μ is the mean of all sample data, and δ is the standard deviation of all sample data,

3.4 Transfer Learning

Transfer learning can apply the trained model to a new but related field, thus the convergence of the model can be completed faster with less training cost in the absence of annotated data. The transfer learning model used in this paper is loaded with the parameters of the Wavegram Logmel CNN [13] trained on Audioset [14] and fine-tuned. After activation function *log softmax* the prediction results of respiratory cycle on four classifications normal, crackle, wheeze and both are obtained. The input original waveform is trained separately on the two branches and the merged result goes through 5 dropout layers and 5 block operations containing *conv2d*, *batchNorm2d*, *relu* and *avg_pool2d*.

More precisely, the branches proceed as follows:

Branch1: After the data is processed by Conv1d with a kernel size of 11 and BatchNorm1d, it goes through three blocks containing *conv1d*, *batchNorm1d*, *relu* and *max_pool1d*.
Branch2: The 8s audio data is subjected to Fourier short-time transform and 64 Mel bins to generate a 701x64 Mel spectrogram. The spectrum is output after a block operation including *conv2d*, *batchNorm2d*, *relu* and *avg_pool2d*.

3.5 Snapshot Ensemble

Snapshot Ensemble [15] is a way to get an ensemble of models from a single training session without additional training cost. It sets the learning rate as shown in Figure 3. Whenever the learning rate restarts, the model starts exploring other local optima and takes model snapshots at different local minima.

Equation 8 shows the mathematical representation of Figure 3, where l_{init} is the initial learning rate, e_i is the i-th epoch in the current cycle, and c is the cycle length. The function f calculates the learning rate for each epoch as follows:

$$f(e_i) = \frac{l_{init}}{2} \cdot (cos(\frac{\pi \cdot i}{c}) + 1) \tag{8}$$

Figure 3 sets the initial learning rate to 0.01, 50 epochs as a cycle, and a total of 4 cycles. Considering the memory problem of a single gpu, the test process can load model snapshots on several gpus and the results obtained are averaged on the same gpu.

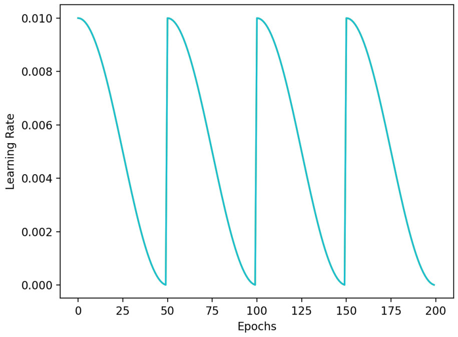


Fig. 3. Learning rate with the current epoch used in the training process

4 Experiments and results

The transfer learning system proposed in this paper is used to solve three tasks: 1) To investigate the effect of different experimental setups on model scores and to compare our work with other works on four classifications of respiratory cycles (normal, crackles, wheezes and both crackles and wheezes); 2) the unbalanced distribution of the dataset in terms of gender, age and BMI was analyzed with demographic information provided by ICBHI and the differences in model scores were validated on the male and female test groups; 3) we proposed a restricted mixup to reduce the differences in model performance between the male and female test groups found in the second task.

4.1 Performance comparison

First, respiratory cycles with a sampling frequency of 16kHz are divided into trainingset and testset according to a ratio of 8:2. We set the baseline of the transfer learning system without using nlpaug library, rollAudio and mixup for data augmentation in the data preprocessing. By filling respiratory cycles with 0 (Zeropadding) or extracting only the first 8s of them we can increase or decrease the length of respiratory cycles to 8s. The baseline uses the SGD optimizer and cross entropy loss.

Hyperparameters We evaluate the scores of baselines using different learning rates and batchsizes on the ICBHI dataset. Referring to the batchsize in [13] we compared its model scores on 16, 32 and 64 (Figure 4). When the batch size is 32, the model score is 2.0% and 2.9% higher than other 2 options. Figure 5 shows that the model scores at learning rate 0.01 are much higher than those at learning rates 0.1 and 0.001. And the model scores in the learning rate based on Equation 8 are higher than the model scores when the learning rate is constant at 0.01. So we will apply a batchsize of 32 and a learning rate based on Equation 8 to the baseline in the following performance comparison.

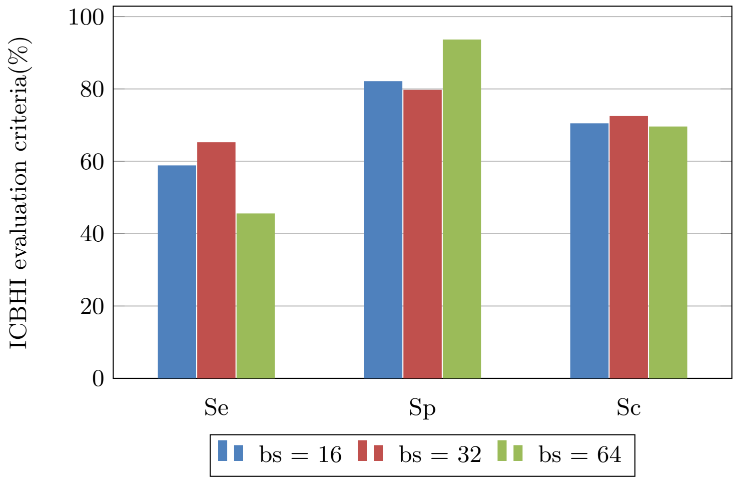


Fig. 4. Comparison of baselines with different batchsize on the testset ($epochs = 50$, $lr = 0.01$)

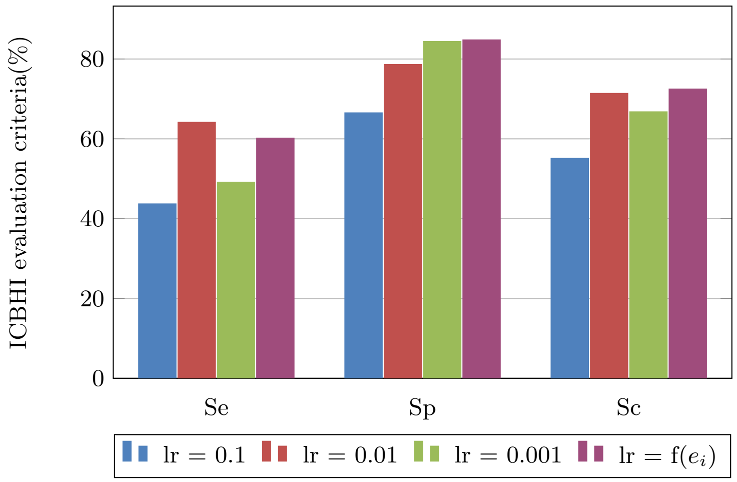


Fig. 5. Comparison of baselines with different learning rates on the testset ($epochs = 50$, $bs = 32$)

Data Preprocessing We focus on the effect of different data preprocessing methods on the model training and loss functions in Figures 6 and 7. Setting the epoch to 200, we add the splitting and padding method to the baseline model identified in the previous subsection, resulting in a 4.3% increase in the model score. The combination of splitting and padding, mixup has only 0.3% difference in model score but the loss function has dropped by 29.2%. We found that using the splitting and padding, mixup and data augmentation combination in the data preprocessing stage resulted in the smallest loss function and the highest model score of 80.9%.

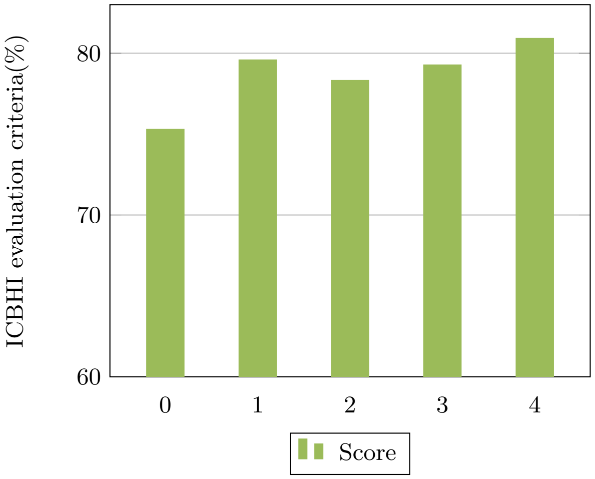


Fig. 6. Comparing model performance with different data preprocessing methods (0: Baseline, 1: SP, 2: SP+rollaudio, 3: SP+mixup, 4: SP+mixup+DA Group) on the test set($epochs = 200$, $bs = 32$)

Snapshot Ensemble From Figures 7 and 8 we see the effectiveness of Snapshot Ensemble in the respiratory cycle 4-classification task. Also, we verify that the data preprocessing scheme with the best performance in the previous subsection still has the highest model score (82.0%) and the smallest loss function value (0.513) after Snapshot Ensemble.

Comparison to other works The most suitable configuration is selected by comparing the performance of our transfer learning system on hyperparameters, data preprocessing and snapshot ensemble in section 4.1. This section compares the model scores of the transfer learning system after a 10-fold cross validation against other work.

The results in table 3 show that our transfer learning system outperforms competitors on all three evaluation criteria. The RNN-based end-to-end model

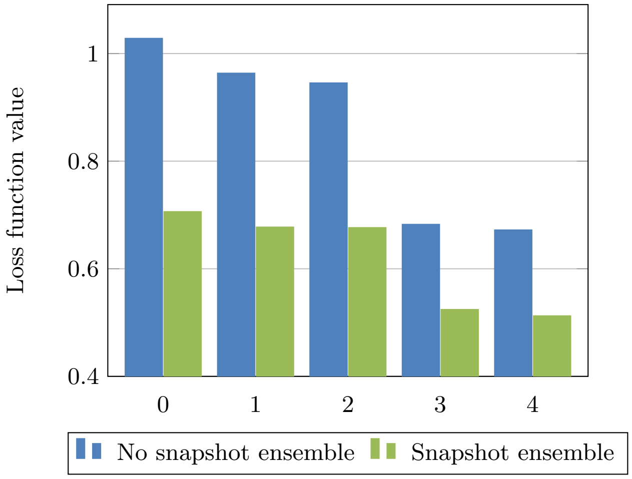


Fig. 7. Compare the loss function values of different data preprocessing methods (0: Baseline, 1: SP, 2: SP+rollaudio, 3: SP+mixup, 4: SP+mixup+DA Group) before and after snapshot ensemble

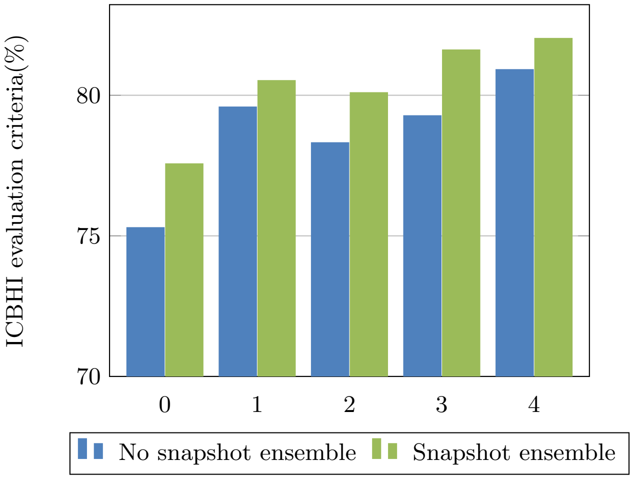


Fig. 8. Compare the model scores of different data preprocessing methods (0: Baseline, 1: SP, 2: SP+rollaudio, 3: SP+mixup, 4: SP+mixup+DA Group) before and after snapshot ensemble

Table 3. ICBHI Challenge Comparison (on four categories)

Method	Se(%)	Sp(%)	Sc(%)
NMRNN [7]	56.0	73.6	64.8
RespireNet [17]	53.7	83.3	68.5
STFT+Wavelet [18]	55.3	83.3	69.3
Hybrid CNN-RNN [8]	56.9	86.7	71.8
LSTM [6]	62.0	85.0	74.0
MBTCNSE [19]	65.3	86.1	75.7
CNN(snapshot ensemble)[20]	69.4	87.3	78.4
Our System	**70.5**	**91.7**	**81.1**

architecture adopted in [7] detects abnormal sounds in the respiratory cycle through masking of noise. [17] adopted the data processing method of device specific fine-tuning, concatenation-based augmentation, blank region clipping, and smart padding and realized the respiratory cycle classification task based on a simple CNN. [18] used STFT and wavelet to extract features and input them into the support vector machine. [6] established a learning framework based on recurrent neural networks to discover time-dependent patterns from sound data. Both [8, 19] completed training on a hybrid neural network: the former added a Bi-LSTM layer for learning temporal features to CNN, and the latter integrated multi-branch temporal convolutional network and squeeze-and-excitation network. [20], which is the closest to us in terms of experimental methods and experimental results, also uses the combined method of CNN and snapshot ensemble, but since our transfer learning system uses more diverse data augmentation methods in data processing (such as random combination of augmentation methods and mixup), using pre-trained network parameters on the large-scale sound dataset Google audioset [14]. Our transfer learning system achieves a score of 81.1% on the test set when compared with other state-of-the-art systems, which is obviously more advantageous.

4.2 Demographic Data

In this section, we analyze the imbalanced distribution of the respiratory cycle dataset based on demographic data on the binary classification task (normal and abnormal). Referring to the labels in Table 1, the normal class includes all normal respiratory cycles, and the abnormal class includes crackle, wheeze and both respiratory cycles. The demographic data provided by ICBHI [2] for each participant were arranged in order of age, gender, BMI, child weight and child height. Considering that there are different classification criteria for children's BMI [21, 22], we excluded respiratory cycles belonging to children and only analyzed the imbalance of the ICBHI dataset on the demographic data age, gender and BMI of adults. We also excluded information on some subjects due to missing age, gender or BMI. After the above two rounds of filtering the dataset contains 6004 respiratory cycles

Table 4. Distribution and overall percentage of subjects' gender on binary classification task (normal and abnormal) corresponding to the respiratory cycle dataset

Normal		Abnormal	
Female	Male	Female	Male
929	2018	1065	1992
15.47%	33.61%	17.74%	33.18%

Table 4 summarizes the data distribution of the respiratory cycle dataset based on different genders when performing the binary classification task. In the normal and abnormal data distributions, males have 18.14% and 15.44% higher than females in the dataset, respectively.

Table 5. Distribution and overall percentage of subjects' age on binary classification task (normal and abnormal) corresponding to the respiratory cycle dataset

Normal		Abnormal	
Adult	Senior	Adult	Senior
500	2447	448	2609
8.33%	40.76%	7.46%	43.45%

We classified subjects aged greater than or equal to 18 years but less than 60 years and those aged greater than or equal to 60 years as adults and seniors, respectively. In Table 5, it is shown that in the respiratory cycle judged as normal, the data for the senior were 4.89 times higher than the data for adults; in the respiratory cycle judged as abnormal, the data for the senior were 5.82 times higher than the data for adults.

Table 6. Distribution and overall percentage of subjects' BMI on binary classification task (normal and abnormal) corresponding to the respiratory cycle dataset

Normal				Abnormal			
Under	Normal	Over	Obesity	Under	Normal	Over	Obesity
135	767	1393	652	596	728	1143	590
2.25%	12.77%	23.20%	10.86%	9.93%	12.13%	19.04%	9.83%

According to the BMI we divided the subjects into underweight ($BMI < 18.5$), normalweight ($18.5 \leq BMI < 25$), overweight ($25 \leq BMI < 30$) and obesity ($BMI \geq 30$). Table 6 shows that the underweight and overweight groups had the smallest (2.25%) and the largest (23.20%) data distribution in the normal category; the obesity and overweight groups had the smallest (9.83%) and the largest (19.04%) data distribution in the abnormal category, respectively.

From tables 4-6, we can see that the subgroups based on demographic data age, gender and BMI, respectively, are unevenly distributed in the binary-classes dataset. Table 7 is based on the summary of tables 4-6. The dataset containing 6004 respiratory cycles can be divided into 32 different attributes based on the three demographic variables gender, age, BMI, and respiratory cycle labels. In studying the differences between female and male transfer learning system, we divided the data corresponding to each attribute into the training-, validation-, and testsets sequentially in the ratio of 72:8:20 to avoid the uneven distribution of the dataset aggravated by the random division.

Table 7. In the binary classification task (normal and abnormal), the uneven distribution of the respiratory cycle dataset on the three demographic variables of gender, age and BMI, each row corresponds to underweight, normalweight, overweight and obesity.

	Normal				Abnormal			
	Female		Male		Female		Male	
	Adult	Senior	Adult	Senior	Adult	Senior	Adult	Senior
underweight	7	80	0	48	0	427	0	169
normalweight	225	115	76	351	238	37	105	348
overweight	16	259	56	1062	15	142	2	984
obesity	0	227	120	305	0	206	88	296

4.3 Comparison of model performance based on subject gender

In this section, three experiments will be conducted based on no mixup, using global mixup and using restricted Mixup, so as to compare the difference in scores of the transfer learning system on the female test group and the male test group (Figure 9) and to ameliorate the discriminatory effect on a particular gender due to training the model on an unbalanced dataset.

In the first experiment, the transfer learning system inherits the best configuration of the model from section 4.1, but does not use mixup for data augmentation in the data preprocessing phase. As known from Figure 9, the model scored 4.44% lower on the male test group than on the female test group. Thus, respiratory cycles from males are less likely to be correctly classified than those from females.

Based on the findings of the first experiment we used mixup in the global scope to perform online data augmentation. Referring to Equation 5,6, we randomly selected respiratory cycle x_j across the entire dataset and fused it with the original respiratory cycle x_i in a ratio of $(1-\lambda)$ to λ, where λ is from the beta distribution. The transfer learning system with the addition of global Mixup increased the score on the male testset by 0.86%, but still scored 3.53% lower than the female testset.

Considering the positive impact of the second experiment, we continued mixup's method in the third experiment but added the restriction. For $G_1, G_2 \in (female, male)$ there are the following formulas,

$$\tilde{x} = \lambda x_{G_{1,i}} + (1 - \lambda) x_{G_{2,j}} \tag{9}$$

$$\tilde{y} = \lambda y_{G_{1,i}} + (1 - \lambda) y_{G_{2,j}} \tag{10}$$

$x_{G_{1,i}}$ and $y_{G_{1,i}}$ are the input and target of the original respiratory cycle from the subject with gender G_1; the input and target of the random cycle for mixup are $x_{G_{2,j}}$ and $y_{G_{2,j}}$ from the subject with gender G_2. The two subjects differ in gender $(G_1 \neq G_2)$. The results returned by Equations (9, 10) will be used as the input to the neural network.

The transfer learning system with the addition of gender discrimination-aware mixup scored 2.6% higher on the male testset than in the first experiment, and the difference in scores with the female testset was reduced to 0.82%. This shows that adding the restriction to mixup can effectively reduce the difference in model performance on the male and female testsets due to the unbalanced dataset and reduce the discriminatory effect of the model on the male testset.

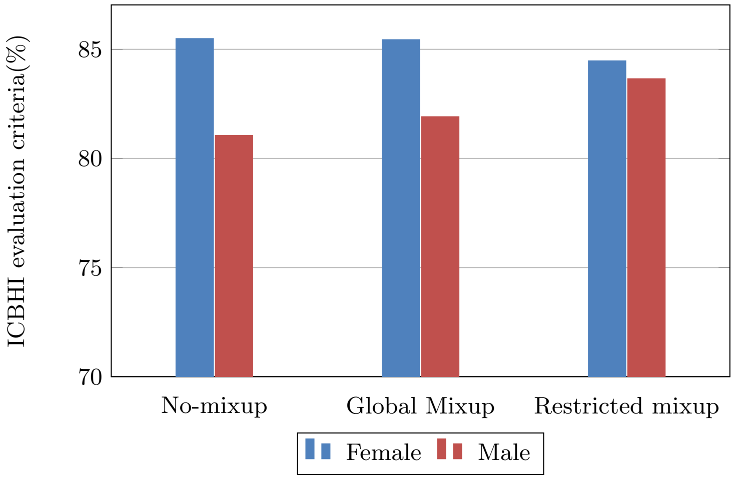


Fig. 9. Comparison of the transfer learning system scores on the female test group and the male test group under different mixup settings

5 Conclusion and future work

Our proposed transfer learning system implements the respiratory cycle 4-classification task (normal, crackle, wheeze or both) in the Respiratory Sound database ICBHI [2] and achieves a score of 81.1% using the evaluation method provided by ICBHI Challenge. After comparing the performance of the transfer learning system based on the pre-trained model Wavegram Logmel CNN [13] with different hyperparameters, data preprocessing methods and model ensembling, we selected the best model configuration that outperformed almost all state-of-the-art systems. In addition, we also discuss the uneven distribution of the respiratory cycle dataset on the gender, age and BMI of the subjects according to the 2-classification task (normal or abnormal respiratory cycle) and validate the 4.44% difference in their scores on the female testset compared to the male testset. For the discriminatory effect of the transfer learning system on the male testset, we explored different mixup methods in the data preprocessing phase and proposed a restricted mixup method to reduce this difference to 0.82%.

In the future work, how to utilize the demographic information corresponding to the respiratory cycle will be the focus of our research. Therefore, the following questions are worth further exploration in this study. (1) Considering the inclusion of subject demographic information (gender, age, and BMI) in the training workflow, how will these information affect the model performance; (2) Due to the uneven distribution of demographic information (gender, age, and BMI) in the respiratory cycle dataset (Table 7), whether the model performance also varies across age and BMI groups. What is the effect of the proposed restricted mixup method in the third experiment on the model performance trained on different age or BMI groups; (3) Whether the proposed restricted mixup method can improve the classification accuracy of the entire dataset.

References

1. World Health Organization. World health statistics 2020[J]. 2020.
2. Rocha B M, Filos D, Mendes L, et al: A respiratory sound database for the development of automated classification, International Conference on Biomedical and Health Informatics. Springer, Singapore, 2017: 33-37.
3. Chambres G, Hanna P, Desainte-Catherine M: Automatic detection of patient with respiratory diseases using lung sound analysis, 2018 International Conference on Content-Based Multimedia Indexing (CBMI). IEEE, 2018: 1-6.
4. Ma Y, Xu X, Yu Q, et al.: LungBRN: A smart digital stethoscope for detecting respiratory disease using bi-resnet deep learning algorithm, 2019 IEEE Biomedical Circuits and Systems Conference (BioCAS). IEEE, 2019: 1-4.
5. Demir F, Sengur A, Bajaj V.: Convolutional neural networks based efficient approach for classification of lung diseases. Health information science and systems, 2020, 8(1): 1-8.
6. Perna D, Tagarelli A.: Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks, 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS). IEEE, 2019: 50-55.

7. Kochetov K, Putin E, Balashov M, et al.: Noise masking recurrent neural network for respiratory sound classification, International Conference on Artificial Neural Networks. Springer, Cham, 2018: 208-217.
8. Acharya J, Basu A.: Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning. IEEE transactions on biomedical circuits and systems, 2020, 14(3): 535-544.
9. Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge[J]. International journal of computer vision, 2015, 115(3): 211-252.
10. Nguyen T, Pernkopf F. Lung Sound Classification Using Co-tuning and Stochastic Normalization[J]. IEEE Transactions on Biomedical Engineering, 2022.
11. Edward Ma.: NLP Augmentation. https://github.com/makcedward/nlpaug
12. Zhang H, Cisse M, Dauphin Y N, et al.: mixup: Beyond empirical risk minimization[J]. arXiv preprint arXiv:1710.09412, 2017.
13. Kong Q, Cao Y, Iqbal T, et al.: Panns: Large-scale pretrained audio neural networks for audio pattern recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020, 28: 2880-2894.
14. Gemmeke J F, Ellis D P W, Freedman D, et al.: Audio set: An ontology and human-labeled dataset for audio events, 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2017: 776-780.
15. Huang G, Li Y, Pleiss G, et al.: Snapshot ensembles: Train 1, get m for free[J]. arXiv preprint arXiv:1704.00109, 2017.
16. McFee B, Raffel C, Liang D, et al. librosa: Audio and music signal analysis in python, Proceedings of the 14th python in science conference. 2015, 8: 18-25.
17. Gairola S, Tom F, Kwatra N, et al. 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). IEEE, 2021: 527-530.
18. Petmezas G, Cheimariotis G A, Stefanopoulos L, et al. Automated Lung Sound Classification Using a Hybrid CNN-LSTM Network and Focal Loss Function. Sensors, 2022, 22(3): 1232.
19. Zhao Z, Gong Z, Niu M, et al. Automatic Respiratory Sound Classification Via Multi-Branch Temporal Convolutional Network. ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022: 9102-9106.
20. Nguyen T, Pernkopf F. Lung sound classification using snapshot ensemble of convolutional neural networks, 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020: 760-763.
21. Rolland-Cachera M F, European Childhood Obesity Group. Childhood obesity: current definitions and recommendations for their use. International Journal of Pediatric Obesity, 2011, 6(5-6): 325-331.
22. Niederer I, Kriemler S, Zahner L, et al. BMI group-related differences in physical fitness and physical activity in preschool-age children: a cross-sectional analysis. Research quarterly for exercise and sport, 2012, 83(1): 12-19.