Interdisciplinary Journal of Engineering and Environmental Sciences
ISSN: 2994-2780 | Impact Factor : 5.76
Volume. 9, Number 1; January-March, 2022;
Published By: Scientific and Academic Development Institute (SADI)
8933 Willis Ave Los Angeles, California
https://sadijournals.org/index.php/ijees/index|editorial@sadijournals.com

LUNG SOUND RECOGNITION BASED ON PRE-TRAINED CONVOLUTIONAL NEURAL NETWORK

Shanshu Bao[1], Lei Liu[2], Bo Che[2] and Linhong Deng[2]

[1]School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, Jiangsu, 213164, China

[2]Changzhou Key Laboratory of Respiratory Medical, Engineering, Institute of Biomedical Engineering and Health Sciences, Changzhou University, Changzhou, Jiangsu, 213164, China

Abstract: It has become one of the focuses of respiratory medicine to recognize lung sounds by machine learning methods, and then assist doctors to diagnose patients' pulmonary diseases. Aiming at the problems of model overfitting and low classification accuracy caused by the small size of lung sound dataset in current lung sound recognition, a lung sound recognition method based on the combination of pre-trained convolutional neural network and CatBoost algorithm was proposed. The pre-trained convolutional neural network on the image dataset ImageNet is transferred to lung sound recognition. The channel attention mechanism CBAM is fused to enhance the recognition performance of the network, and the lung sound waveform data is converted into logarithmic MEL frequency spectrum for input and training. Finally, the trained model is used as a feature extractor, and the feature vectors with high-level semantics are input into the ensemble learning algorithm CatBoost to achieve the final classification. After experiment, the result shows that the specificity, sensitivity and ICHBI-score of the proposed method for lung sound recognition in ICBHI-2017 lung sound dataset reach 88.34%, 63.13% and 75.73%, respectively, which is superior to the previous methods. The display has a good application prospect in lung sound recognition.

Keywords: lung sound recognition; convolutional neural network; channel attention; ensemble learning

1. Introduction

Lung sound signal is an extremely important physiological signal in clinical auscultation, which contains a lot of physiological information and plays an important role in the diagnosis and monitoring of human health. At present, in clinical auscultation, doctors mainly rely on subjective experience to judge the type of lung sounds of patients to assist their diagnosis of lung diseases. However, only relying on the subjective experience of doctors may lead to missed diagnosis or even misdiagnosis of pulmonary diseases. Chest x-rays, lung function tests and other X-ray methods, although widely used in clinical practice, are harmful to the human body. In view of the fact that lung sound auscultation does not cause additional harm to human body, with the development of

computer science, it has become an important research trend to use machine learning methods to realize the recognition and classification of lung sound signals, and then assist doctors to diagnose patients' pulmonary diseases.

Abnormal lung sounds do not exist independently, but are added to the normal lung sounds. If there are no crackles, wheezes, or wheezes in a lung sound signal, it is considered as a normal lung sound. The main differences between the three types of abnormal lung sounds are their frequency and duration. A crackle is about 1200Hz and lasts 5 to 15 milliseconds; a wheeze is 400Hz or more and lasts 80 to 100 milliseconds; and a wheeze is less than 300Hz and lasts more than 100 milliseconds. The key to lung sound recognition includes the extraction and characterization of lung sound signal features, among which Mel Frequency spectrum (MS) and Mel Frequency Cepstral Coefficient (MFCC) are commonly used as signal features, and other acoustic signal representation methods to characterize the energy of sound signals in each frequency range, the envelope of sound formants and other important feature information.

However, there are two main problems facing the research of lung sound recognition: (1) the existing recognition algorithms are difficult to obtain key information from the features of lung sound; (2) Lack of publicly available large data sets. The most widely used lung sound dataset in recent years is the open lung sound dataset ICBHI-2017 from the International Conference on Biomedical Informatics (ICBHI) [1], which contains 6898 annotated four-class lung sound data from 126 subjects. Based on the study of

-70-

lung sound recognition in this dataset, Serbes G et al. [2] adopted the method of feature fusion. After fusing the short-time Fourier spectrum features and wavelet features of lung sounds, support vector machine (SVM) was used to identify four types of lung sounds in the ICBHI dataset, with an ICBHIscore of 67.29%. Demir F[3] et al. adopted the short-time Fourier spectrum features of lung sounds and used a convolutional neural network named parallel-pooling CNN to classify and identify four types of lung sounds, with an ICBHI-score of only 70.45%. Ma Y[4] et al. extracted the short-time Fourier spectrum features of lung sounds and proposed a convolutional neural network named Bi-ResNet to identify lung sounds, with ICBHI-score of only 69.30%. It can be seen that the existing methods are difficult to extract key information from the characteristics of lung sounds. Therefore, it is urgent to develop a lung sound recognition method based on data self-drive, high recognition accuracy and excellent comprehensive performance.

In conclusion, considering that the pre-trained model used on large datasets can overcome the problem of low model recognition accuracy caused by the small size of the dataset, the ensemble learning method performs well in various competition datasets. Therefore, with the help of the pre-trained model on ImageNet, this study proposes a method based on pre-trained convolutional neural network combined with CatBoost[5] to recognize lung sounds. Our main contributions are as follows: we use time translation to expand the training data, and fuse the channel attention CBAM [6] module in the pre-training model to enhance the ability of recognition network. The trained model is treated as a feature extractor, and the feature vectors with high-level semantics are fused into ensemble learning CatBoost algorithm to realize classification.

2. Method for lung sound recognition

2.1. Pre-trained convolutional neural network

ImageNet is the most authoritative image dataset. In order to use the pre-training parameters of convolutional network on ImageNet, we need to convert the waveform data of lung sounds into the image form of 224×224×3 . The main steps are as follows:

(1): the waveform data of lung sounds is converted to the spectrogram.

(2): Make three copies of the spectrogram in (1) and join the spectrogram.

There are many forms of spectrogram. Considering the excellent performance of logarithmic MEL frequency spectrum in audio recognition and classification, we converted lung sound signal to logarithmic MEL frequency spectrum. FIG. 1 shows the time-domain waveform of the lung sound data used in the experiment and the logarithmic MEL frequency spectrum transformed in this way.

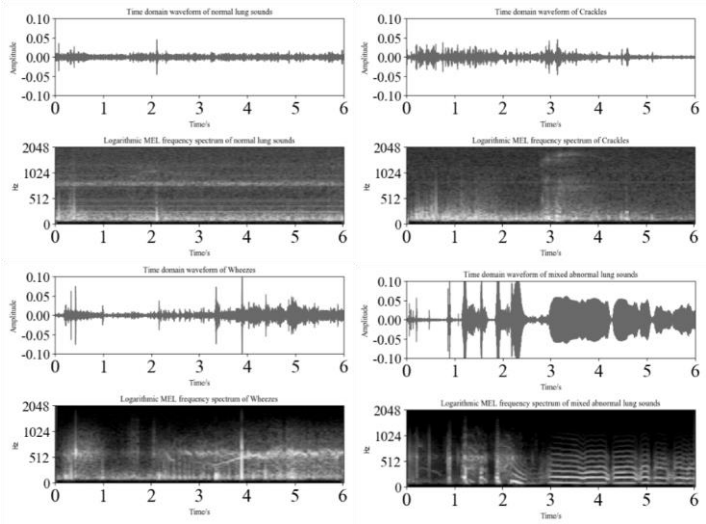

Figure 1: Time domain diagram and logarithmic MEL frequency spectrum of different types of lung sounds

2.2. CBAM

To improve the recognition effect of CNN, we fuse the channel attention module CBAM in the network structure. The Convolutional Block Attention Module (CBAM) is an implementation of the channel attention mechanism. The core function of the module is as follows: by learning, the module allocates weights to the number of channels of the feature vector, so as to facilitate the attention to important features and the suppression of minor features. Figure 2 shows the structure of CBAM.

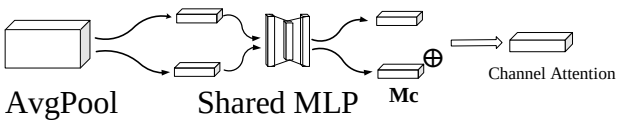

Figure 2: Schematic diagram of CBAM structure

Let the data dimension of the input feature vector F be$W×H×C$. After *MaxPool* and *AvgPool*, two $1×1×C$ feature vectors are obtained. These vectors are added and summed by the shared perceptron, and the channel attention map M_C is obtained by the function. M_C is the weight of the corresponding channel. The number

on each channel is multiplied by the corresponding weight of the channel, and the weight distribution of each channel is completed. In particular, the calculation process of M_C can be given by formula (1).

$$M_C(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (1)$$

In formula (1), *MaxPool* and *AvgPool* represent maximum pooling and average pooling, respectively, *MLP* represents multi-layer perceptron, and σ represents *sigmoid* function.

2.3. Method introduction

After converting the waveform data of lung sounds into logarithmic MEL frequency spectrum, the pre-trained convolutional neural network with channel attention is input, and the feature vectors with high-level semantics are input into CatBoost algorithm for classification. The specific schematic diagram is given in Figure 3.

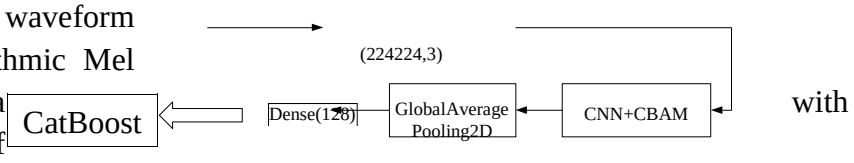

Figure 3: Schematic diagram of lung sound recognition based on CNN.

It is worth mentioning that CatBoost and neural network cannot be combined for one training. Therefore, convolutional neural network is trained with training data first, and then convolutional neural network is used as feature extractor to extract high-level semantic vectors, and 128-dimensional feature vectors are used to train CatBoost. CatBoost as one of the mainstream algorithms for gradient Boosting decision trees, can reduce overfitting by using the Ordering Boosting algorithm.

3. Experiment data

3.1. Dataset

The open dataset provided by the International Conference on Biomedical Health Informatics 2017 (ICBHI-2017) was used as the experimental dataset in this study. The dataset contains 6898 annotated 4category lung sounds data from 126 subjects, including 3642 normal lung sounds, 1864 crackle sounds, 886 wheeze sounds, and 506 abnormal lung sounds mixed with crackle and wheeze sounds. The data set is divided into training set and test set according to 8:2. The sampling time of lung sound data is set to 6 seconds.

3.2. Evaluation indicators

The ICBHI lung sound experiment dataset in this study has special evaluation indicators: Specificity (Sp), Sensitivity (Se) and ICBHI-score, and the relevant calculation formula is shown as follows.

$$Sp = \frac{TP^{Normal}}{TP_{Normal} + FN_{Normal}} \quad (2)$$

$$Se = \frac{TP^{Abnormal}}{TP_{Abnormal} + FN_{Abnormal}} \quad (3)$$

$$ICBHI\text{-}score = \frac{Sp + Se}{2} \quad (4)$$

Where, TP represents the number of samples whose true value is positive class and which the model also identifies as positive class. FN represents the number of samples whose true value is positive class and the model identifies as negative class. The positive class can be any type of lung sounds, and the other lung sounds except the positive class are the negative class. Normal means healthy lung sounds, Abnormal includes wheezing, crackling, and a mixture of the two.

3.3. Data Augment

Due to the small number of data and uneven distribution in the data set, data augmentation was used to expand the lung sound data with fewer categories. In order to have no impact on the composition of lung sounds, this study used the method of time translation for data augmentation, which is to translate the waveform along the time axis. Figure 4 illustrates this process.

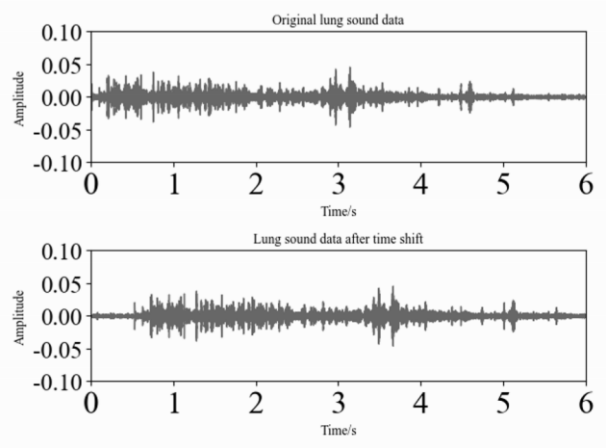

Figure 4: Schematic diagram of data enhancement

After data augmentation, the number of crackles in the training set changed from 1491 to 2982, wheezes from 722 to 2167, and mixed abnormal lung sounds from 405 to 3280. This measure has alleviated the over fitting phenomenon caused by less data.

4. Experimental results and analysis

4.1. Lung sound recognition effect of different pre-trained models on ImageNet

As mentioned above, we realized lung sound recognition with the help of the pre-trained model on ImageNet, which is the most authoritative representative image dataset, and there are various pre-trained models on ImageNet. We tested different pre-trained models on ImageNet on the lung sound dataset, and the results are shown in Table 1.

Table 1: Comparison of recognition effects of different pre-trained models on lung sounds on ImageNet

Pretrained Model	Sp(%)	Se(%)	ICBHI-score
VGG16	86.14	60.06	73.10
VGG19	88.34	63.13	75.73
ResNet50	77.91	49.76	63.83
DenseNet121	77.22	49.61	63.42
InceptionNet	80.93	44.39	62.66
EfficientNet	79.01	50.84	64.92

According to the data in Table 1, compared with other convolutional neural networks, VGG19 has the best recognition effect on lung sounds. In addition, we found that the more complex the structure of the network, the effect is not ideal. This also shows from the side that the existing lung sound data is indeed not enough, resulting in serious overfitting and weak generalization ability of the network. In view of the fact that VGG19 has a better recognition effect on lung sounds than its convolutional neural network, VGG19 is used for subsequent studies. The confusion matrix shown in Figure 5 shows the specific classification of each lung sound by VGG19.

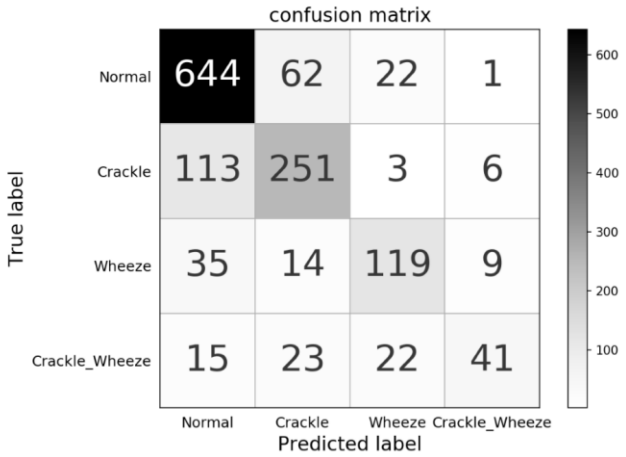

Figure 5: Confusion matrix of lung sound recognition by VGG19

4.2. Influence of channel attention and CatBoost on the model

We added CBAM to VGG19 to increase the recognition performance, and combined with CatBoost algorithm to achieve classification. The following ablation experiments were performed to compare the model after CBAM and CatBoost deletion with VGG19, and the results are shown in Table 2.

Table 2: Ablation experiments

Deleted Module	Sp(%)	Se (%)	ICBHI-score
CBAM	86.55	60.98	73.76
CatBoost	86.00	60.21	73.10
CBAM+CatBoost	85.45	59.90	72.67
None	88.34	63.13	75.73

As shown in Table 2, deleting either CatBoost or CBAM module in VGG19 will reduce the ICBHIscore of the model, and when both modules are deleted at the same time, the model performance will decrease significantly.

4.3. Influence of different channel attention on the model

In this study, CBAM channel attention was added to VGG19 to improve the model recognition ability. However, in addition to CBAM, SE-NET [7] and ECA-NET [8] are also known channel attention mechanisms. The main function of each of these mechanisms is to assign weights to different channels, and the difference is only in how they are implemented. To evaluate the effect of different mechanisms on the performance of the model proposed in this study, we replaced CBAM with SE-NET or ECA-NET respectively, and then compared the recognition performance of the model on lung sounds. The results are shown in Table 3.

Table 3: Effects of different channel attention on the model

Channel Attention	Sp(%)	Se (%)	ICBHI-score
----	86.55	60.98	73.76
SE-Net	85.73	61.13	73.43
ECA-Net	88.20	62.05	75.12
CBAM	88.34	63.13	75.73

The results show that different channel attention mechanisms can increase the ICBHI-score of the model, but compared with SE-NET and ECA-NET, CBAM has a greater effect on the performance improvement of the model, which indicates the correctness of the idea of integrating CBAM in VGG19. ***4.4. The influence of different classifiers on the model***

Instead of using the FC layer directly after the neural network to realize the classification of lung sounds, we used the trained VGG19 as a feature extractor and input the extracted feature vectors into CatBoost for training and classification. It can not be ignored that XGBoost[9] and LightGBM[10] are similar to CatBoost. They all belong to the mainstream algorithms in gradient lifting decision trees and play significant roles in some researches. In addition, the combination of neural network and support vector machine (SVM) has also achieved excellent results in the field of some small datasets. Therefore, we experimentally compare the influence of different classifiers on the recognition effect, and results are shown in Table 4.

Table 4: Effects of different classifiers on the model

Type of classifier	Sp(%)	Se (%)	ICBHI-score
FC layer	86.00	60.21	73.10
SVM	87.24	60.52	73.88
XGBoost	87.51	61.13	74.32
LightGBM	87.10	61.59	74.34
CatBoost	88.34	63.13	75.73

According to the data in the above table, no matter using SVM or combined with Boosting algorithm, the recognition effect obtained is better than directly using fully connected layer. Since CatBoost had the highest ICBHI-score, we finally chose it as a classifier.

4.5. The influence of pre-training parameters on the model

As mentioned above, pre-training parameters of convolutional neural network on ImageNet are used in this study. In order to verify the importance of pre-training parameters, we conduct comparative experiments, and results are shown in Table 5.

Table 5: Influence of pre-training parameters on model performance

Use pre-trained parameters	Sp(%)	Se (%)	ICBHI-score
No	84.08	39.93	62.00
Yes	88.34	63.13	75.73

According to the data in the table, the pre-training parameters play a very important role. Once the pre-training parameters are not used, the performance of the model decreases significantly. This also shows from the side that the existing lung sound data is indeed rare, it is difficult to train a better neural network.

4.6. Comparison of recognition effects of different recognition methods on ICBHI-2017

Table6 shows the comparison of the results of the method in this study and the method proposed by predecessors in the lung sound type recognition experiment on the ICBHI-2017. It can be seen that the ICBHI-score obtained by the model proposed in this study is higher than other models.

Table 6: Comparison of recognition effects of different recognition methods on ICBHI-2017

Type of classifier	Feature	Sp(%)	Se (%)	ICBHI-score
SVM[2]	STFT+ Wavelet	83.25	55.29	69.27
Parallel-Pooling CNN[3]	STFT	83.24	57.67	70.45
Bi-ResNet[4]	STFT	80.06	58.54	69.30
CNN[11]	Spectro-gram	83.00	53.00	68.00
ResNet-FC +DataAugment[12]	LogMel	83.30	53.70	68.50
CNN-CatBoost	LogMel	88.34	63.13	75.73

As shown in Table 6, the Sp, Se and ICBHI-score obtained by the CNN-CatBoost proposed in this paper are higher than other methods proposed by predecessors. In addition, it is not difficult to find that when the data augmentation or transfer learning method is not used, the obtained recognition effect is not good, which again verifies the correctness of the study starting from the transfer learning method.

5. Conclusions

This study proposes a CNN-CatBoost model for lung sound recognition. We draw support from the pre-training parameters of VGG19 on ImageNet to overcome the overfitting phenomenon caused by less data to a certain extent, and integrate the recognition performance of CBAM channel attention enhancement network, and finally use CatBoost for classification. The experimental results on the ICBHI-2017 show that the ICBHI-score of the model reaches 75.73%, which is superior to the existing lung sound recognition methods using CNN.

In addition, the following questions are worth further exploration in this study :(1) when lung sound data and computer power allow, how to use lung sound data training parameters instead of pre-training parameters to obtain a recognition model more suitable for the acoustic characteristics of lung sounds, so as to provide a higher precision auxiliary recognition technology for clinical lung sound auscultation?

References

Rocha B M, Filos D, Mendes L, et al. A respiratory sound database for the development of automated classification[C].International Conference on Biomedical and Health Informatics. Springer, Singapore, 2017: 33-37.

Serbes G, Ulukaya S, Kahya Y P. An automated lung sound preprocessing and classification system based onspectral analysis methods[C].International Conference on Biomedical and Health Informatics. Springer, Singapore, 2017: 45-49.

Demir F, Ismael A M, Sengur A. Classification of lung sounds with CNN model using parallel pooling structure [J]. IEEE Access, 2020, 8: 105376-105383.

Ma Y, Xu X, Yu Q, et al. LungBRN: A smart digital stethoscope for detecting respiratory disease using bi-resnet deep learning algorithm[C].2019 IEEE Biomedical Circuits and Systems Conference (BioCAS). IEEE, 2019: 1-4.

Prokhorenkova L, Gusev G, Vorobev A, et al. CatBoost: unbiased boosting with categorical features [J]. Advances in Neural Information Processing Systems, 2018, 31.

Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C].Proceedings of the European Conference on Computer Vision (ECCV). 2018: 3-19.

Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7132-7141.

Wang Q, Wu B et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks [C].2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ:IEEE Press, 2020.

Chen T, Guestrin C. Xgboost: A scalable tree boosting system[C].Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016: 785-794.

Ke G, Meng Q, Finley T, et al. Lightgbm: A highly efficient gradient boosting decision tree [J]. Advances in neural information processing systems, 2017, 30.

Demir F, Sengur A, Bajaj V. Convolutional neural networks based efficient approach for classification of lung diseases [J]. Health Information Science and Systems, 2020, 8(1): 1-8.

Gairola S, Tom F, Kwatra N, et al. Respirenet: A deep neural network for accurately detecting abnormal lung sounds in limited data setting[C].2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2021: 527-530.